

Optimal Copying of One Quantum Bit

Chi-Sheng Niu* and Robert B. Griffiths†
 Department of Physics, Carnegie Mellon University,
 Pittsburgh, PA 15213

A quantum copying machine producing two (in general non-identical) copies of an arbitrary input state of a two-dimensional Hilbert space (qubit) is studied using a quality measure based on distinguishability of states, rather than fidelity. The problem of producing optimal copies is investigated with the help of a Bloch sphere representation, and shown to have a well-defined solution, including cases in which the two copies have unequal quality, or the quality depends upon the input state (is “anisotropic” in Bloch sphere language), or both. A simple quantum circuit yields the optimal copying machine. With a suitable choice of parameters it becomes an optimal eavesdropping machine for some versions of quantum cryptography, or reproduces the Bužek and Hillery result for isotropic copies.

I. INTRODUCTION

The no-cloning theorem [1,2] is one of the most important features which distinguishes quantum from classical information theories. There is no difficulty making an arbitrary number of copies of any type of information which arrives over a classical channel. However, to copy or “clone” the information which arrives over a quantum channel is not possible without producing errors if it is encoded in terms of nonorthogonal quantum states. This fact is the basis for various schemes of quantum cryptography [3], since the attempt of an eavesdropper to tap into a quantum channel and duplicate the information in it will result in errors detectable by the legitimate users.

Even in cases in which perfect copies are excluded because of the no-cloning theorem, it is possible to produce imperfect copies which are better than random noise. An optimal copying machine is one in which, by means of a suitable unitary transformation, a certain number of copies of the original information are produced with the smallest number of errors, or minimal amount of noise. Optimality in this sense will obviously depend upon the ensemble of input states, the number of copies, and the measure employed for noise or error rate. Obtaining quantitative estimates of what is possible for optimal copying is a significant problem in quantum information theory, and it is to this problem that the present paper is addressed.

A number of previous studies have used various input ensembles and measures of quality or of noise. Bužek and Hillery discussed an “isotropic”, or “universal”, copying machine [4], which produces two identical copies from one qubit (two-state system), with copy quality independent of the input state. This copying machine was later proved by Bruß *et al.* [5] to be optimal among all possible isotropic machines, if the measure of quality is the fidelity between the input and the output: that is, the probability that the output will be measured to be in the same state as the input. Gisin and Massar [6] considered a case in which M identical copies are generated from N identical qubits ($M > N$), and found that the average fidelity reaches its maximum when the machine is isotropic. Bruß, Ekert, and Macchiavello [7] found an interesting connection between the optimal isotropic copying process and quantum state estimation by measurement. Later, the optimal M -to- N problem was generalized by Werner [8] to systems of arbitrary dimension. Various quantum circuits implementing some of the isotropic copying machines mentioned above are presented in [9,10]. Bruß *et al.* [5] also considered an input ensemble of two non-orthogonal states, and found that the optimal copying process requires no ancillary qubits, unlike the case of isotropic copying.

In contrast to the work just discussed, we will use a quality measure based upon the distinguishability of the output states, rather than the fidelity, and will consider *non-identical* and *anisotropic* copies for which the copying quality can depend upon the input state. This is motivated by a study of eavesdropping in

*Electronic mail: cn28+@andrew.cmu.edu

†Electronic mail: rgrif@cmu.edu

quantum cryptography [11]. Eavesdropping is, in essence, a matter of producing imperfect “copies” of the input signal in some channel in a way which enables the eavesdropper to distinguish the different states in the original information, while at the same time perturbing it as little as possible. Thus distinguishability is more important to the eavesdropper than whether his copies resemble the original states. In addition, the amount of information obtained by the legitimate receiver and by the eavesdropper are generally different, and can depend on the input states. Since the problem of optimizing a copy machine for completely general inputs and outputs is quite difficult, we restrict ourselves to the case of only two copies of one qubit. Even with this restriction, the mathematical structure is nontrivial. Recently Cerf [12] explored this problem by investigating a family of “Pauli cloning machines.” He derived a no-cloning inequality for isotropic copies, and established a bound for the quantum capacity of the corresponding “Pauli channel.”

We introduce the background material needed for later sections in Sec. II, including a formalism for the general interaction of one qubit with another quantum system, a set of superoperators, and the Bloch sphere representation. In Sec. III, the *quality function* of an output qubit is defined, and it is expressed using the Bloch sphere representation; this is the measure we employ to define optimal copying. Given the quality of one copy, the best quality for the other is determined by a map G , which is derived in Sec. IV for a special case. However, this special case is as good or better than any other copying scheme, as shown in Sec. V. Section VI discusses a number of properties of the map G , the conditions under which a pair of copies produced are actually optimal, and particular examples that are useful in revealing relations between the copying qualities. In Sec. VII we introduce two quantum circuits implementing the optimal copying process, and show how they can be turned into an eavesdropping machine or a universal copying machine. Finally, the results are summarized and some open questions are noted in Sec. VIII. Various mathematical details are placed in Appendices A–D.

II. GENERAL FORMULATION OF THE COPYING PROBLEM

We define the copying problem in the following way. The qubit which is to be copied is a state in a two-dimensional Hilbert space \mathcal{A} with orthogonal basis $|0\rangle$ and $|1\rangle$. The copy machine corresponds to a unitary transformation U on

$$\mathcal{H} = \mathcal{A} \otimes \mathcal{E} = \mathcal{B} \otimes \mathcal{E} = \mathcal{B} \otimes \mathcal{C} \otimes \mathcal{D}, \quad (2.1)$$

where we suppose that the copies lie in the two dimensional spaces \mathcal{B} and \mathcal{C} after the action of U , and the space $\mathcal{E} = \mathcal{C} \otimes \mathcal{D}$ can have an arbitrarily large but finite (even) dimension. The identical spaces \mathcal{A} and \mathcal{B} are denoted differently in order to distinguish the input space from the output. In particular,

$$U(|j\rangle \otimes |\epsilon\rangle) = \sum_{l=0}^1 |l\rangle \otimes |\epsilon_l^j\rangle, \quad (2.2)$$

where $|\epsilon\rangle$ is a normalized state of \mathcal{E} , and therefore the states $|\epsilon_l^j\rangle$ of \mathcal{E} satisfy

$$\sum_l \langle \epsilon_l^j | \epsilon_l^k \rangle = \delta_{jk}. \quad (2.3)$$

There is no loss of generality in assuming a pure state $|\epsilon\rangle$ rather than a density matrix for \mathcal{E} , since a density matrix can always be thought of as arising from a partial trace on a space of higher dimension, and the dimension of \mathcal{E} is arbitrary.

With $|\epsilon\rangle$ fixed, define the linear map $V : \mathcal{A} \rightarrow \mathcal{B} \otimes \mathcal{E}$ through

$$V|\alpha\rangle = U(|\alpha\rangle \otimes |\epsilon\rangle) = \sum_{l=0}^3 (\sigma_l |\alpha\rangle) \otimes |\epsilon_l\rangle, \quad (2.4)$$

where σ_0 is the identity, and σ_1, σ_2 , and σ_3 are the usual Pauli matrices σ_x, σ_y , and σ_z , which in (2.4) should be thought of as maps from \mathcal{A} to \mathcal{B} , assuming some basis $|0\rangle, |1\rangle$ for the latter; here $|0\rangle$ and $|1\rangle$ correspond to S_z equal to $+1/2$ and $-1/2$, respectively, and $(|0\rangle + |1\rangle)/\sqrt{2}$ to $S_x = +1/2$ in the usual spin half notation. Thus if (2.4) is written out explicitly, it takes the form:

$$\begin{aligned} V|0\rangle &= |0\rangle|e_0\rangle + |1\rangle|e_1\rangle + i|1\rangle|e_2\rangle + |0\rangle|e_3\rangle, \\ V|1\rangle &= |1\rangle|e_0\rangle + |0\rangle|e_1\rangle - i|0\rangle|e_2\rangle - |1\rangle|e_3\rangle, \end{aligned} \quad (2.5)$$

where, following the usual convention, we omit \otimes if this causes no confusion. The four $|e_l\rangle$ are vectors in \mathcal{E} which are linear combinations of the $|e_k^j\rangle$, and whose inner products form a 4×4 positive (i.e., non-negative eigenvalues) Hermitian matrix E :

$$E_{lm} = \langle e_l | e_m \rangle = E_{ml}^*. \quad (2.6)$$

Note that given a positive Hermitian matrix E , it is always possible to find a set of four vectors $|e_l\rangle$ such that E_{lm} takes the form (2.6), and this set is unique up to an arbitrary unitary transformation on \mathcal{E} .

The condition (2.3) implied by the unitarity of U is equivalent to the requirement that V be an isometry, or that

$$V^\dagger V = \sum_{lm} E_{lm} \sigma_l \sigma_m = \sigma_0, \quad (2.7)$$

the identity on \mathcal{A} , and thus to

$$\sum_{l=0}^3 E_{ll} = 1, \quad (2.8)$$

$$\text{Re}(E_{0q}) = \text{Im}(E_{q'q''}), \quad (2.9)$$

where Re and Im refer to real and imaginary parts, and (2.9) employs a notation which we will use again and again:

$$(q, q', q'') = (1, 2, 3) \text{ or } (2, 3, 1) \text{ or } (3, 1, 2). \quad (2.10)$$

That is, the triple (q, q', q'') is an *even* permutation of the integers 1, 2, and 3. Thus, for example, $\text{Re}[E_{02}] = \text{Im}[E_{31}]$.

Given V , the corresponding *superoperator* \hat{V} which maps operators on \mathcal{A} to operators on \mathcal{H} takes the form

$$\hat{V}(\kappa) = V \kappa V^\dagger = \sum_{jk} \sigma_j \kappa \sigma_k \otimes |e_j\rangle\langle e_k|. \quad (2.11)$$

In particular, for $l = 0, 1, 2, 3$,

$$\hat{V}(\sigma_l) = \sum_{jkm} L(jk; lm) \sigma_m \otimes |e_j\rangle\langle e_k| = \sum_m \sigma_m \otimes \hat{F}_{lm}, \quad (2.12)$$

where the \hat{F}_{lm} are a collection of 16 Hermitian operators on \mathcal{E} defined by

$$\hat{F}_{lm} = \sum_{jk} L(jk; lm) |e_j\rangle\langle e_k|, \quad (2.13)$$

and

$$L(jk; lm) = \frac{1}{2} \text{Tr}[\sigma_j \sigma_l \sigma_k \sigma_m] \quad (2.14)$$

can be thought of as a 16×16 matrix, with rows labeled by jk and columns by lm . Its properties are discussed in App. A. In particular, $\frac{1}{2}L$ is its own inverse. When $j = k$ or when $l = m$, $L(jk; lm)$ vanishes unless the other pair is also equal:

$$L(jk; ll) = \delta_{jk} \Lambda(j, l), \quad L(jj, lm) = \delta_{lm} \Lambda(j, l), \quad (2.15)$$

where $\Lambda(j, l)$ is a 4×4 real matrix, (A8), whose square is 4 times the identity.

The superoperators \hat{V}_B , \hat{V}_C , and \hat{V}_E , which play a central role in the following discussion, map operators on \mathcal{A} to operators on \mathcal{B} , \mathcal{C} , and \mathcal{E} , respectively and are defined using partial traces of \hat{V} :

$$\begin{aligned}\hat{V}_{\mathcal{B}}(\kappa) &= \text{Tr}_{\mathcal{E}}[\hat{V}(\kappa)], & \hat{V}_{\mathcal{E}}(\kappa) &= \text{Tr}_{\mathcal{B}}[\hat{V}(\kappa)], \\ \hat{V}_{\mathcal{C}}(\kappa) &= \text{Tr}_{\mathcal{B} \otimes \mathcal{D}}[\hat{V}(\kappa)] = \text{Tr}_{\mathcal{D}}[\hat{V}_{\mathcal{E}}(\kappa)]\end{aligned}\quad (2.16)$$

The isometry condition (2.7) implies that \hat{V} preserves the trace,

$$\text{Tr}_{\mathcal{H}}[\hat{V}(\kappa)] = \text{Tr}_{\mathcal{A}}[\kappa], \quad (2.17)$$

and this property is inherited by each of the superoperators $\hat{V}_{\mathcal{B}}$, $\hat{V}_{\mathcal{C}}$, and $\hat{V}_{\mathcal{E}}$, with the trace over \mathcal{H} in (2.17) replaced by a trace over \mathcal{B} , \mathcal{C} , and \mathcal{E} , respectively. (Henceforth we shall not always indicate explicitly the space over which a trace is to be taken if it is obvious from the context.) In addition, it is obvious from (2.11) that \hat{V} maps Hermitian operators to Hermitian operators, and this property is also inherited by $\hat{V}_{\mathcal{B}}$, $\hat{V}_{\mathcal{C}}$, and $\hat{V}_{\mathcal{E}}$.

It is convenient to write

$$\hat{V}_{\mathcal{B}}(\sigma_l) = \sum_m B_{lm} \sigma_m, \quad (2.18)$$

where

$$B_{lm} = \text{Tr}_{\mathcal{E}}[\hat{F}_{lm}] = \sum_{jk} L(lm; jk) E_{jk} \quad (2.19)$$

is a real 4×4 matrix of the form:

$$B = \begin{pmatrix} 1 & \delta_1^B & \delta_2^B & \delta_3^B \\ 0 & & & \\ 0 & & \mathbf{B} & \\ 0 & & & \end{pmatrix}, \quad (2.20)$$

where columns and rows are in the order 0, 1, 2, 3, and \mathbf{B} is a real 3×3 matrix. The form of the first (left-most) column of B is equivalent to the isometry condition (2.7), that is, to (2.8) plus (2.9), as can be shown by inserting in (2.19) the matrix elements for the L given in App. A. It implies that the superoperator in (2.18) preserves the trace. The inverse of (2.19), see (A10), is

$$E_{jk} = \frac{1}{4} \sum_{lm} L(jk; lm) B_{lm}. \quad (2.21)$$

The vector $\boldsymbol{\delta}^B = (\delta_1^B, \delta_2^B, \delta_3^B)$ and the matrix \mathbf{B} in (2.20) can be given a geometrical interpretation using the well-known Bloch sphere representation

$$\rho = \frac{1}{2}(I + \mathbf{r} \cdot \boldsymbol{\sigma}) = \frac{1}{2}(\sigma_0 + \sum_{j=1}^3 r_j \sigma_j) \quad (2.22)$$

of a density matrix ρ in a two-dimensional Hilbert space, where $\mathbf{r} = (r_1, r_2, r_3)$ is any real vector of length less than 1 for a mixed state, or equal to 1 for a pure state, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. Then (2.18) and (2.20) tell us that

$$\hat{V}_{\mathcal{B}}(\rho) = \rho' = \frac{1}{2}(\sigma_0 + \mathbf{s} \cdot \boldsymbol{\sigma}), \quad (2.23)$$

where [13]

$$\mathbf{s} = \boldsymbol{\delta}^B + \mathbf{r} \cdot \mathbf{B} = \boldsymbol{\delta}^B + \mathbf{B}^T \cdot \mathbf{r}, \quad (2.24)$$

and \mathbf{B}^T is the transpose of \mathbf{B} . In particular, if (2.24) is applied to the collection of pure states, $|\mathbf{r}| = 1$, which form the unit Bloch sphere of \mathcal{A} , the resulting collection of \mathbf{s} vectors forms an ellipsoid in the Bloch sphere of \mathcal{B} , which we shall call the \mathcal{B} ellipsoid, or simply ellipsoid, with center at $\boldsymbol{\delta}^B$.

The polar decomposition [14] of \mathbf{B} takes the form

$$\mathbf{B} = \mathbf{P} \cdot \mathbf{B}^s, \quad (2.25)$$

where \mathbf{P} is a proper rotation, a real orthogonal 3×3 matrix with determinant 1, and \mathbf{B}^s is a real symmetric matrix. By diagonalizing \mathbf{B}^s , one can always write \mathbf{B} in the form

$$\mathbf{B} = \mathbf{P}' \cdot \mathbf{B}^d \cdot \mathbf{P}'', \quad (2.26)$$

where \mathbf{P}' and \mathbf{P}'' are proper rotations, and \mathbf{B}^d is a diagonal matrix with (real) eigenvalues b_1, b_2, b_3 . The rotations \mathbf{P}' and \mathbf{P}'' correspond to unitary transformations on \mathcal{A} and \mathcal{B} , respectively. Consequently, by adopting suitable orthonormal bases for \mathcal{A} and \mathcal{B} , we can, assuming U to be given, always arrange to have the matrix B in the form

$$B = \begin{pmatrix} 1 & \delta_1^B & \delta_2^B & \delta_3^B \\ 0 & b_1 & 0 & 0 \\ 0 & 0 & b_2 & 0 \\ 0 & 0 & 0 & b_3 \end{pmatrix}, \quad (2.27)$$

parameterized by $\delta^B = (\delta_1^B, \delta_2^B, \delta_3^B)$ and $\mathbf{b} = (b_1, b_2, b_3)$. (Note that in the process of diagonalizing the general form (2.20) to obtain the diagonal form (2.27), the values of the δ_j^B will, in general, change.)

Two vectors corresponding to an orthonormal basis or *mode* for \mathcal{A} are represented by opposite poles $\hat{\mathbf{m}} = (m_1, m_2, m_3)$ and $-\hat{\mathbf{m}}$ of the Bloch sphere of \mathcal{A} , where we use a hat to indicate a unit vector. Under (2.24) these map to opposite points of the \mathcal{B} ellipsoid; that is, the line connecting them passes through the center δ^B of the ellipsoid. The three principal axes of the \mathcal{B} ellipsoid correspond to three *principal modes* of \mathcal{A} , represented in the Bloch sphere of \mathcal{A} by mutually perpendicular unit vectors (and their negations) $\hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2$, and $\hat{\mathbf{M}}_3$. If bases are chosen such that B has the form (2.27), the mode vectors $\hat{\mathbf{M}}_j$ are along the Cartesian axes, and correspond, in the language of spin half, to a value of $+1/2$ for S_x, S_y , and S_z , respectively. (It is, of course, important to distinguish “orthogonal” with reference to the complex Hilbert spaces \mathcal{A} or \mathcal{B} , from “perpendicular” as it refers to real three-dimensional vectors in the Bloch sphere representation!) Also, $|b_j|$ is the length of the j 'th principal semi-axis of the ellipsoid.

Because \mathbf{P}' and \mathbf{P}'' are proper rotations (each has determinant 1), it is not always possible to make all the b_j , $j = 1, 2, 3$, in (2.27) positive (non-negative). If an even number of the b_j are negative, then \mathbf{P}' and \mathbf{P}'' can be chosen so they are all positive, but if an odd number are negative, the best one can do is to have two positive and one negative. Which of these situations occurs depends upon U , and turns out to be of some significance for optimal copying.

Since \mathcal{C} is also a two-dimensional space, we can use a Bloch sphere representation for $\hat{V}_{\mathcal{C}}$ and write the counterpart of (2.23) as

$$\hat{V}_{\mathcal{C}}(\rho) = \rho'' = \frac{1}{2}(\sigma_0 + \mathbf{t} \cdot \boldsymbol{\sigma}), \quad (2.28)$$

where

$$\mathbf{t} = \delta^C + \mathbf{r} \cdot \mathbf{C} = \delta^C + \mathbf{C}^T \cdot \mathbf{r}. \quad (2.29)$$

The counterpart of (2.20) is a 4×4 matrix

$$C = \begin{pmatrix} 1 & \delta_1^C & \delta_2^C & \delta_3^C \\ 0 & & & \\ 0 & & \mathbf{C} & \\ 0 & & & \end{pmatrix}, \quad (2.30)$$

where \mathbf{C} is a 3×3 matrix which can be diagonalized by the same strategy employed earlier in the case of \mathbf{B} , though the principal modes of \mathcal{A} singled out by this construction will, in general, not be the same as those corresponding to the \mathcal{B} ellipsoid, so we label the perpendicular unit vectors with a prime: $\hat{\mathbf{M}}'_1, \hat{\mathbf{M}}'_2$, and $\hat{\mathbf{M}}'_3$. We shall denote by $\mathbf{c} = (c_1, c_2, c_3)$ the values on the diagonal when \mathbf{C} is diagonalized; hence $|c_j|$ is the length of the j 'th principal semi-axis of the \mathcal{C} ellipsoid.

The possible values of \mathbf{b} and δ^B are constrained by the fact that, whatever U may be, the matrix E , which determines B through (2.19), and is determined by B through the inverse relationship (2.21), satisfies the isometry condition (2.8) and (2.9) and has non-negative eigenvalues. One consequence is the fact that the \mathcal{B} ellipsoid must always lie inside the unit Bloch sphere of \mathcal{B} . These constraints are found explicitly for the case $\delta^B = \mathbf{0}$ in Sec. IV below, but we do not know their form in general. There are, of course, identical constraints on the possible values of \mathbf{c} and δ^C .

III. ERROR RATES AND QUALITY FUNCTION

Consider the problem of a general channel of the type introduced in Sec. II, from a qubit \mathcal{A} to a Hilbert space \mathcal{F} (which could be \mathcal{B} or \mathcal{C} or \mathcal{E}), described by a superoperator $\hat{V}_{\mathcal{F}}$. Suppose that symbols a_1 and a_2 are encoded in basis states $\hat{\mathbf{m}}$ and $-\hat{\mathbf{m}}$ in the Bloch sphere representation of \mathcal{A} , and the corresponding states are mapped by $\hat{V}_{\mathcal{F}}$ to the two density matrices $\rho_{\mathcal{F}}^1$ and $\rho_{\mathcal{F}}^2$ in \mathcal{F} . Now suppose that every time a signal a_1 or a_2 is sent, a measurement is carried out on \mathcal{F} corresponding to a decomposition of the identity

$$I_{\mathcal{F}} = \sum_k F_k, \quad (3.1)$$

where the $\{F_k\}$ are projectors, or, more generally, positive operators [15]. Further suppose that if the result of this measurement is f_k corresponding to F_k , an estimate of the original symbol is made according to a fixed guessing function¹ $g(a_j|f_k)$ which is either 0 or 1, and satisfies

$$\sum_j g(a_j|f_k) = 1. \quad (3.2)$$

That is, if $g(a_2|f_1) = 1$, then if f_1 is measured, one guesses that a_2 was sent. If a_j is sent with probability p_j , one can show that the average error rate (guessing a_1 when a_2 was sent, or vice versa) using this procedure is bounded below, see [16], by the (achievable) minimum error rate

$$\epsilon_{min} = \frac{1}{2} \left[1 - \text{Tr} \left(|p_1 \rho_{\mathcal{F}}^1 - p_2 \rho_{\mathcal{F}}^2| \right) \right], \quad (3.3)$$

where the absolute value $|C|$ of a Hermitian operator C is the operator obtained by replacing the eigenvalues of C with their absolute values in its spectral decomposition. Motivated by (3.3), we define the *quality factor* $\mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{m}})$ or “distinguishability” for mode $\hat{\mathbf{m}}$ as

$$\mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{m}}) = \frac{1}{2} \text{Tr} \left(|\rho_{\mathcal{F}}^1 - \rho_{\mathcal{F}}^2| \right) = \text{Tr} \left(|\Omega_{\mathcal{F}}(\hat{\mathbf{m}})| \right), \quad (3.4)$$

where

$$\Omega_{\mathcal{F}}(\hat{\mathbf{m}}) = \hat{V}_{\mathcal{F}} \left(\frac{1}{2} \hat{\mathbf{m}} \cdot \boldsymbol{\sigma} \right) = \frac{1}{2} \hat{\mathbf{m}} \cdot \hat{V}_{\mathcal{F}}(\boldsymbol{\sigma}). \quad (3.5)$$

Thus $\mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{m}})$ is a number between 0 and 1, equal to $1 - 2\epsilon_{min}$ in the case in which $p_1 = p_2 = 1/2$. A quality factor of 1 means the two signals can be perfectly distinguished, while 0 means that measurements are no more effective than random guesses.

The quality factor as a function of mode, $\mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{m}})$, will be called the *quality function*. Knowledge of this function provides a general measure of “distinguishability” for pairs of non-orthogonal as well as orthogonal signals in \mathcal{A} , for if a_1 and a_2 correspond to density matrices

$$a_j \leftrightarrow \rho_{\mathcal{A}}^j = \frac{1}{2} (I + \mathbf{x}^j \cdot \boldsymbol{\sigma}) \quad (3.6)$$

in \mathcal{A} , the distinguishability of the corresponding signals in \mathcal{F} , compare (3.4), is

$$\frac{1}{2} \text{Tr} \left(|\hat{V}_{\mathcal{F}}(\rho_{\mathcal{A}}^1) - \hat{V}_{\mathcal{F}}(\rho_{\mathcal{A}}^2)| \right) = \frac{1}{4} x \text{Tr} \left(|\mathbf{x} \cdot \hat{V}_{\mathcal{F}}(\boldsymbol{\sigma})| \right) = \frac{1}{2} x \mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{x}}), \quad (3.7)$$

where $\hat{\mathbf{x}}$ is a unit vector in the direction of $\mathbf{x}^1 - \mathbf{x}^2$, and $x = |\mathbf{x}^1 - \mathbf{x}^2|$.

The quality factor $\mathcal{Q}_{\mathcal{F}}(\hat{\mathbf{m}})$ for mode $\hat{\mathbf{m}}$ is, (3.4), the sum of the absolute values of the eigenvalues of $\Omega_{\mathcal{F}}(\hat{\mathbf{m}})$. When $\mathcal{F} = \mathcal{H}$, $\hat{V}_{\mathcal{F}} = \hat{V}$, and since \hat{V} is an isometry, $\hat{V}(\frac{1}{2} \hat{\mathbf{m}} \cdot \boldsymbol{\sigma})$ has two non-zero eigenvalues, $+1/2$

¹One can employ values for $g(a_j|f_k)$ between 0 and 1, provided (3.2) is satisfied, for a random guessing process in which $g(a_j|f_k)$ is the probability of guessing a_j given f_k . However, one can show that this more general procedure does not result in a smaller error rate, so we shall not consider it further.

and $-1/2$, the same as those of $\frac{1}{2}\hat{\mathbf{m}} \cdot \boldsymbol{\sigma}$. Hence $\mathcal{Q}_{\mathcal{H}}(\hat{\mathbf{m}}) = 1$, independent of $\hat{\mathbf{m}}$. When $\mathcal{F} = \mathcal{B}$ or \mathcal{E} , it is helpful to note that $\Omega_{\mathcal{B}}$ and $\Omega_{\mathcal{E}}$ can be obtained from partial traces of

$$\Omega_{\mathcal{H}}(\hat{\mathbf{m}}) = \frac{1}{2} \sum_{q=1}^3 m_q \sum_{l=0}^3 \sigma_l \otimes \hat{F}_{ql}, \quad (3.8)$$

see (2.12). In particular, one has

$$\Omega_{\mathcal{B}}(\hat{\mathbf{m}}) = \text{Tr}_{\mathcal{E}}[\Omega_{\mathcal{H}}(\hat{\mathbf{m}})] = \frac{1}{2}\hat{\mathbf{m}} \cdot \mathbf{B} \cdot \boldsymbol{\sigma}, \quad (3.9)$$

since B_{lm} , (2.19), is the trace of \hat{F}_{lm} , and $B_{q0} = 0$, see (2.20). The eigenvalues of $\Omega_{\mathcal{B}}(\hat{\mathbf{m}})$ are $\pm \frac{1}{2}|\hat{\mathbf{m}} \cdot \mathbf{B}|$, where $\hat{\mathbf{m}} \cdot \mathbf{B}$ (see the discussion in Sec. II) is the vector from the center of the ellipsoid to the point on its surface which is the image, under $\hat{V}_{\mathcal{B}}$, of $\hat{\mathbf{m}}$, and $|\hat{\mathbf{m}} \cdot \mathbf{B}|$ is its length. The same argument applies when \mathcal{F} is \mathcal{C} , and thus we have

$$\mathcal{Q}_{\mathcal{B}}(\hat{\mathbf{m}}) = |\hat{\mathbf{m}} \cdot \mathbf{B}|, \quad \mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}}) = |\hat{\mathbf{m}} \cdot \mathbf{C}|. \quad (3.10)$$

If we write $\hat{\mathbf{m}}$ in terms of the principal modes of \mathcal{A} relative to the \mathcal{B} and \mathcal{C} ellipsoids,

$$\hat{\mathbf{m}} = \sum_{j=1}^3 \mu_j \hat{\mathbf{M}}_j = \sum_{j=1}^3 \mu'_j \hat{\mathbf{M}}'_j, \quad (3.11)$$

respectively, the two quality functions can be written explicitly as:

$$\mathcal{Q}_{\mathcal{B}}(\hat{\mathbf{m}}) = \sqrt{\left\{ \sum_{j=1}^3 (b_j \mu_j)^2 \right\}}, \quad \mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}}) = \sqrt{\left\{ \sum_{j=1}^3 (c_j \mu'_j)^2 \right\}}, \quad (3.12)$$

where the b_j and the c_j are the diagonal elements of the diagonalized matrices \mathbf{B} and \mathbf{C} , and represent in each case the length of the j 'th principal semi-axis of the ellipsoid.

In the case $\mathcal{F} = \mathcal{E}$, we have

$$\Omega_{\mathcal{E}}(\hat{\mathbf{m}}) = \sum_{q=1}^3 m_q \hat{F}_{q0} \quad (3.13)$$

upon tracing (3.8) over \mathcal{B} . There is no simple expression (known to us) for the spectrum of $\Omega_{\mathcal{E}}(\hat{\mathbf{m}})$, and we cannot represent the quality function $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ using a Bloch sphere. Nonetheless, it is worth noting that $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ is completely determined if the superoperator $\hat{V}_{\mathcal{B}}$, or, equivalently, the matrix \mathbf{B} , is given. That is because \mathbf{B} determines \mathbf{E} , (2.21), and as noted below Eq. (2.6), it thus determines \hat{F}_{lm} , see (2.13), up to a unitary transformation on \mathcal{E} . As a consequence, this transformation cannot change $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ since it leaves the spectrum of $\Omega_{\mathcal{E}}(\hat{\mathbf{m}})$ the same. On the other hand, $\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}})$ is not uniquely determined by $\hat{V}_{\mathcal{B}}$, because there are many ways in which \mathcal{E} can be expressed as a tensor product $\mathcal{C} \otimes \mathcal{D}$. Nonetheless, it is clear that

$$\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}}) \leq \mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}}), \quad (3.14)$$

because measurements on \mathcal{C} alone clearly cannot do better at distinguishing signals than those carried out on the full space \mathcal{E} . A formal proof of (3.14) is obtained by noting that

$$\Omega_{\mathcal{C}}(\hat{\mathbf{m}}) = \text{Tr}_{\mathcal{D}}[\Omega_{\mathcal{E}}(\hat{\mathbf{m}})], \quad (3.15)$$

and using theorem 3 of App. B.

IV. OPTIMAL COPYING: CENTERED ELLIPSOID

We approach the problem of optimal copying in the following way. Suppose a copying machine produces one copy in the \mathcal{B} output channel characterized by a quality function $\mathcal{Q}_{\mathcal{B}}(\hat{\mathbf{m}})$ which, as noted in Sec. III,

is completely specified by a set of principal modes in \mathcal{A} together with the semi-axes of the corresponding ellipsoid in the \mathcal{B} Bloch sphere, the absolute values of b_1 , b_2 , and b_3 . We then ask, given $\mathbf{b} = (b_1, b_2, b_3)$, what is the best possible copy which can be produced in a separate one qubit (two-dimensional) output channel \mathcal{C} ; that is, what is the best or largest quality function $\mathcal{Q}_C(\hat{\mathbf{m}})$?

It is not obvious at the outset that this question has a well-defined answer, for one can certainly imagine that with a fixed \mathbf{b} , one type of copy machine might produce an optimal quality in \mathcal{C} for mode $\hat{\mathbf{m}}$, and a different machine would be needed to make the best copies for a different mode $\hat{\mathbf{m}}'$. But we will show that for a given \mathbf{b} there is, indeed, a single machine which produces the best possible copies in \mathcal{C} , as measured by \mathcal{Q}_C , for every $\hat{\mathbf{m}}$. That is, given \mathbf{b} , there is a largest ellipsoid in the Bloch sphere of \mathcal{C} , characterized by semi-axes c_1 , c_2 , and c_3 , with

$$\mathbf{c} = (c_1, c_2, c_3) = G(\mathbf{b}), \quad (4.1)$$

a well-defined function of \mathbf{b} . The construction, incidentally, yields non-negative c_j .

In the present section, the optimization map G will be derived by considering the special case of an ellipsoid centered in the Bloch sphere of \mathcal{B} , that is, $\delta^B = 0$ in (2.27). That considering this special case is actually sufficient will be shown in Sec. V, where we use a concavity property of the quality function to show that the quality cannot be further improved by using an ellipsoid which is *not* centered in the Bloch sphere of \mathcal{B} . However, given the pair \mathbf{b} and $\mathbf{c} = G(\mathbf{b})$, it is not at all obvious that the corresponding copy machine is optimal, since there might be a machine which produces better copies both in the \mathcal{B} channel and in the \mathcal{C} channel. This possibility is studied below in Sec. VI, where we also work out certain properties of G .

If the $|e_l\rangle$ in (2.4) are mutually orthogonal, the matrix E_{lm} of inner products, (2.6) is diagonal, and can be written in the form

$$E_{lm} = \beta_l^2 \delta_{lm}, \quad (4.2)$$

where the β_l are non-negative numbers. Further, we can choose an orthonormal basis $\{|\hat{e}_l\rangle\}$ in \mathcal{E} such that

$$|e_l\rangle = \beta_l |\hat{e}_l\rangle, \quad (4.3)$$

The symbol β without a subscript will be used to denote the four vector $(\beta_0, \beta_1, \beta_2, \beta_3)$. Because of (2.15), if E is diagonal, the matrix B , (2.19), is also diagonal,

$$B_{lm} = b_l \delta_{lm}, \quad (4.4)$$

and we will use b without a subscript for the four vector of diagonal elements:

$$b = (b_0, b_1, b_2, b_3) = (1, \mathbf{b}). \quad (4.5)$$

Using (2.19), (2.15), and (4.2), we can express the elements of b in terms of the β_j through

$$b = \Lambda \cdot \beta^2, \quad \beta^2 = \frac{1}{4} \Lambda \cdot b, \quad (4.6)$$

where the dot denotes matrix multiplication, and

$$\beta^2 = (\beta_0^2, \beta_1^2, \beta_2^2, \beta_3^2) \quad (4.7)$$

is a four vector whose components are the squares of those of β . If (4.6) is written out explicitly one finds:

$$b_0 = 1 = \sum_{j=0}^3 \beta_j^2 \quad (4.8)$$

$$b_q = \beta_0^2 + \beta_q^2 - \beta_{q'}^2 - \beta_{q''}^2 \text{ for } q = 1, 2, 3. \quad (4.9)$$

Since the off-diagonal δ^B in (2.20) is zero, (4.4), the ellipsoid is centered, and the fact that B is diagonal means that the principal axes of the ellipsoid coincide with the Cartesian axes of the Bloch sphere. Note that the property last mentioned can always be achieved by choosing an appropriate orthonormal basis in \mathcal{B} , so that there is no loss of generality in assuming, if the ellipsoid is centered, that the B matrix is diagonal, of the form (4.4). But in that case (2.21)—note (2.15)—tells us that the matrix E is diagonal, so that the

$|e_l\rangle$ must be mutually orthogonal. Thus by considering all mutually orthogonal sets of $|e_l\rangle$, we take care of all cases in which the \mathcal{B} ellipsoid is centered. The fact that each component of β^2 in (4.6) is non-negative, together with $b_0 = 1$, places a set of four constraints on the components of \mathbf{b} :

$$\sum_{p=1}^3 b_p \geq -1; \quad b_q + b_{q'} \leq 1 + b_{q''} \text{ for } q = 1, 2, 3. \quad (4.10)$$

These are the necessary and sufficient conditions for \mathbf{b} to be physically possible when $\delta^B = 0$, and they specify that it lies within a tetrahedron with vertices

$$(1, -1, -1), (-1, 1, -1), (-1, -1, 1), (1, 1, 1). \quad (4.11)$$

As noted at the end of Sec. III, \mathbf{b} in (4.9) determines $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ but not $\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}})$, which depends on the tensor product structure $\mathcal{C} \otimes \mathcal{D}$ of \mathcal{E} . One choice that maximizes $\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}})$ is the set of basis vectors $|\hat{e}_l\rangle$ with

$$\begin{aligned} \sqrt{2}|\hat{e}_0\rangle &= |00\rangle + |11\rangle, & \sqrt{2}|\hat{e}_1\rangle &= |01\rangle + |10\rangle, \\ \sqrt{2}|\hat{e}_2\rangle &= -i|01\rangle + i|10\rangle, & \sqrt{2}|\hat{e}_3\rangle &= |00\rangle - |11\rangle, \end{aligned} \quad (4.12)$$

where orthonormal basis vectors for the two-dimensional space \mathcal{C} are denoted by $|0\rangle$ and $|1\rangle$, the same notation is used for the basis of \mathcal{D} , and in the kets on the right sides in (4.12) the \mathcal{C} label is to the left of the \mathcal{D} label, thus $|cd\rangle$. The space \mathcal{D} is two-dimensional, but could be higher-dimensional with additional basis vectors $|2\rangle, |3\rangle, \dots$ etc., which do not appear in (4.12). Substituting (4.12) into (4.3), and (4.3) into (2.5), we obtain the explicit expressions

$$\begin{aligned} \sqrt{2}V|0\rangle &= (\beta_0 + \beta_3)|000\rangle + (\beta_0 - \beta_3)|011\rangle + (\beta_1 + \beta_2)|101\rangle + (\beta_1 - \beta_2)|110\rangle, \\ \sqrt{2}V|1\rangle &= (\beta_0 - \beta_3)|100\rangle + (\beta_0 + \beta_3)|111\rangle + (\beta_1 - \beta_2)|001\rangle + (\beta_1 + \beta_2)|010\rangle, \end{aligned} \quad (4.13)$$

where the labels in the kets on the right side are in the order $\mathcal{B}, \mathcal{C}, \mathcal{D}$, that is to say, $|bcd\rangle$. Note that in Bloch sphere notation, (2.22), $|0\rangle$ corresponds to $\mathbf{r} = (0, 0, 1)$, $|1\rangle$ to $\mathbf{r} = (0, 0, -1)$, $(|0\rangle + |1\rangle)/\sqrt{2}$ to $\mathbf{r} = (1, 0, 0)$, etc. Keeping this correspondence in mind, it is easy to check that the superoperator $\hat{V}_{\mathcal{B}}$ based upon (4.13) maps $\mathbf{r} = (1, 0, 0), (0, 1, 0)$, and $(0, 0, 1)$ to $\mathbf{s} = (b_1, 0, 0), (0, b_2, 0)$, and $(0, 0, b_3)$, respectively, where the b_q are given by (4.9).

We now define a four vector $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ in terms of β through:

$$\gamma = \frac{1}{2}\Lambda \cdot \beta, \quad \beta = \frac{1}{2}\Lambda \cdot \gamma, \quad (4.14)$$

and rewrite (4.13) in terms of the γ_j as:

$$\begin{aligned} \sqrt{2}V|0\rangle &= (\gamma_0 + \gamma_3)|000\rangle + (\gamma_1 + \gamma_2)|011\rangle + (\gamma_0 - \gamma_3)|101\rangle + (\gamma_1 - \gamma_2)|110\rangle, \\ \sqrt{2}V|1\rangle &= (\gamma_1 + \gamma_2)|100\rangle + (\gamma_0 + \gamma_3)|111\rangle + (\gamma_1 - \gamma_2)|001\rangle + (\gamma_0 - \gamma_3)|010\rangle. \end{aligned} \quad (4.15)$$

Note that if the first two labels, the \mathcal{B} and \mathcal{C} bits, are interchanged on the right side of (4.13) and, at the same time, each β_j is replaced by γ_j , the result is (4.15); that is to say, *the γ_j play precisely the same role for the space \mathcal{C} as the β_j for the space \mathcal{B}* . Consequently, we can immediately conclude that the ellipsoid in the Bloch sphere of the \mathcal{C} channel, produced by the action of $\hat{V}_{\mathcal{C}}$, is centered and has principal semi-axes c_1, c_2 , and c_3 lying parallel to the three Cartesian axes, with

$$c = (c_0, c_1, c_2, c_3) = (1, \mathbf{c}) = \Lambda \cdot \gamma^2, \quad \gamma^2 = \frac{1}{4}\Lambda \cdot c, \quad (4.16)$$

the counterpart of (4.6), and γ^2 the four vector whose j 'th component is γ_j^2 . Inserting (4.14) in (4.16), one obtains explicit expressions for the c_q with $q > 0$:

$$c_q = 2(\beta_0\beta_q + \beta_{q'}\beta_{q''}), \quad (4.17)$$

where q' and q'' follow the notation of (2.10). In addition, since the principal axes of the ellipsoids in the \mathcal{B} and \mathcal{C} Bloch spheres correspond to the same principal modes in \mathcal{A} , we can omit the primes in (3.11) and (3.12). Thus, in particular, the quality function for \mathcal{C} is given by

$$\mathcal{Q}_C(\hat{\mathbf{m}}) = \sqrt{\left\{ \sum_{j=1}^3 (c_j \mu_j)^2 \right\}}. \quad (4.18)$$

The next step is to show that (4.18) is, indeed, the *optimal* quality function, given b_1 , b_2 , and b_3 . To demonstrate that this is the case, we shall show that (3.14) is satisfied as an equality: $\mathcal{Q}_C(\hat{\mathbf{m}}) = \mathcal{Q}_E(\hat{\mathbf{m}})$. In order to evaluate $\mathcal{Q}_E(\hat{\mathbf{m}})$, one needs to find the eigenvalues of the Hermitian operator $\Omega_E(\hat{\mathbf{m}}) = \hat{V}_E(\frac{1}{2}\hat{\mathbf{m}} \cdot \boldsymbol{\sigma})$, whose matrix

$$\langle \hat{e}_l | \Omega_E(\hat{\mathbf{m}}) | \hat{e}_m \rangle = \sum_{j=1}^3 \mu_j L(lm; j0) \beta_l \beta_m, \quad (4.19)$$

when written out explicitly using (3.13) and (2.13), takes the form:

$$\Omega_E = \begin{pmatrix} 0 & \mu_1 \beta_0 \beta_1 & \mu_2 \beta_0 \beta_2 & \mu_3 \beta_0 \beta_3 \\ \mu_1 \beta_0 \beta_1 & 0 & i \mu_3 \beta_1 \beta_2 & -i \mu_2 \beta_1 \beta_3 \\ \mu_2 \beta_0 \beta_2 & -i \mu_3 \beta_1 \beta_2 & 0 & i \mu_1 \beta_2 \beta_3 \\ \mu_3 \beta_0 \beta_3 & i \mu_2 \beta_1 \beta_3 & -i \mu_1 \beta_2 \beta_3 & 0 \end{pmatrix}. \quad (4.20)$$

In (4.19) and (4.20) we have replaced each component m_q of $\hat{\mathbf{m}}$ by the corresponding μ_q , since—see (3.11)—the principal modes \hat{M}_j are along the coordinate axes.

The characteristic polynomial of (4.20) is

$$\det(\Omega_E - \lambda I) = \lambda^4 - \Gamma \lambda^2 + \Delta^2, \quad (4.21)$$

where—once again employing the notation in (2.10)—

$$\Gamma = \sum_{q=1}^3 \mu_q^2 (\beta_0^2 \beta_q^2 + \beta_{q'}^2 \beta_{q''}^2), \quad \Delta = \beta_0 \beta_1 \beta_2 \beta_3 (\sum_{q=1}^3 \mu_q^2). \quad (4.22)$$

Evidently, the eigenvalues of \mathcal{Q}_E occur in pairs $\pm \lambda_1$, $\pm \lambda_2$, where we assume that λ_1 and λ_2 are positive. Thus $\text{Tr}(|\Omega_E|)$ is $2(\lambda_1 + \lambda_2)$, and one finds, after a bit of algebra, that

$$\mathcal{Q}_E(\hat{\mathbf{m}}) = \text{Tr}(|\Omega_E(\hat{\mathbf{m}})|) = \sqrt{\Gamma + 2|\Delta|} = \sqrt{\left\{ \sum_{j=1}^3 (\check{c}_j \mu_j)^2 \right\}} \quad (4.23)$$

where

$$\check{c}_q = 2(|\beta_0 \beta_q| + |\beta_{q'} \beta_{q''}|). \quad (4.24)$$

Absolute value signs have been inserted because even though we have assumed that β_j are non-negative in (4.2), we could use negative values in (4.19) or (4.20), and it is clear from (4.21) and (4.22) that the spectrum of Ω_E depends only on the squares of the β_j . On the other hand, as long as the β_j are non-negative, one sees, by comparing (4.17) with (4.24), and (4.18) with (4.23), that $\mathcal{Q}_C(\hat{\mathbf{m}})$ and $\mathcal{Q}_E(\hat{\mathbf{m}})$ are, indeed, identical. Hence, if the ellipsoid is centered in the \mathcal{B} Bloch sphere, and the values b_1 , b_2 , and b_3 are specified, the construction associated with (4.12) or, equivalently, (4.13), with the β_j assumed to be non-negative, yields copies in the \mathcal{C} channel of the highest possible quality, whatever mode is chosen; see the discussion associated with (3.14).

V. NON-CENTERED CASE: SYMMETRY AND CONCAVITY

A. Centered Case is Optimal

The problem of optimal copying has been solved in Sec. IV in the sense that if the ellipsoid is *centered* in the \mathcal{B} channel and has principal semi-axes $\mathbf{b} = (b_1, b_2, b_3)$, then the optimal quality for copies in the \mathcal{C} channel can be achieved using a centered ellipsoid with principal semi-axes $\mathbf{c} = (c_1, c_2, c_3) = G(\mathbf{b})$. Could

the copies in \mathcal{C} be further improved by employing a copy machine for which the \mathcal{B} ellipsoid is *not* centered in the Bloch sphere? That it cannot follows from the argument given below, which is based upon some properties of $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ which are themselves not without interest, and are derived in Sec. V B and Sec. V C below.

As noted at the end of Sec. III, $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ for a fixed mode $\hat{\mathbf{m}}$ is completely determined by the 4×4 matrix B of (2.20). Since for the following argument we can assume that $\hat{\mathbf{m}}$ is held fixed, it will be convenient to denote the dependence of $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$ on B by $\mathcal{Q}_{\mathcal{E}}[B]$ or, when B has the diagonal form (2.27), by $\mathcal{Q}_{\mathcal{E}}[\mathbf{b}, \boldsymbol{\delta}^B]$, suppressing the reference to $\hat{\mathbf{m}}$. Here \mathbf{b} is the set of ellipsoid semi-axes and $\boldsymbol{\delta}^B$ is the displacement of the ellipsoid from the center of the Bloch sphere.

The first property of $\mathcal{Q}_{\mathcal{E}}[B]$ which is of interest is the symmetry

$$\mathcal{Q}_{\mathcal{E}}[\mathbf{b}, \boldsymbol{\delta}^B] = \mathcal{Q}_{\mathcal{E}}[\mathbf{b}, -\boldsymbol{\delta}^B] \quad (5.1)$$

proved in Sec. V B. The second, proved in Sec. V C, is that $\mathcal{Q}_{\mathcal{E}}[B]$ is a concave function in the sense that, whenever B^1 and B^2 are acceptable B matrices, that is, generated by isometries of the form (2.4), and p_1 and p_2 are positive numbers summing to 1, $p_1 B^1 + p_2 B^2$ is also an acceptable matrix, and

$$\mathcal{Q}_{\mathcal{E}}[p_1 B^1 + p_2 B^2] \geq p_1 \mathcal{Q}_{\mathcal{E}}[B^1] + p_2 \mathcal{Q}_{\mathcal{E}}[B^2]. \quad (5.2)$$

Given the symmetry and concavity properties just mentioned, the fact that an “off center” copying machine cannot improve upon the “centered” version of Sec. IV can be shown in the follow way. Let the copy machine produce an ellipsoid $[\mathbf{b}, \boldsymbol{\delta}^B]$ in the \mathcal{B} channel, and consider a particular mode $\hat{\mathbf{m}}$ for which this machine produces a copy quality $\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}})$ in the \mathcal{C} channel; of course, $\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}})$ cannot, (3.14), exceed $\mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}})$, which we denote by $\mathcal{Q}_{\mathcal{E}}[\mathbf{b}, \boldsymbol{\delta}^B]$. Then, with an obvious choice for B^1 and B^2 , and with $p_1 = p_2 = 1/2$, (5.2) tells us that

$$\mathcal{Q}_{\mathcal{E}}[\mathbf{b}, 0] \geq (\mathcal{Q}_{\mathcal{E}}[\mathbf{b}, \boldsymbol{\delta}^B] + \mathcal{Q}_{\mathcal{E}}[\mathbf{b}, -\boldsymbol{\delta}^B])/2 = \mathcal{Q}_{\mathcal{E}}[\mathbf{b}, \boldsymbol{\delta}^B], \quad (5.3)$$

where we have made use of the symmetry (5.1). In Sec. IV we showed that for a centered ellipsoid $[\mathbf{b}, 0]$ it is possible to construct an optimal copying machine which in the \mathcal{C} channel produces copies of quality

$$\mathcal{Q}_{\mathcal{C}}^{opt}(\hat{\mathbf{m}}) = \mathcal{Q}_{\mathcal{E}}[\mathbf{b}, 0], \quad (5.4)$$

and, consequently, the \mathcal{C} quality for the “off center” machine is bounded by

$$\mathcal{Q}_{\mathcal{C}}(\hat{\mathbf{m}}) \leq \mathcal{Q}_{\mathcal{C}}^{opt}(\hat{\mathbf{m}}). \quad (5.5)$$

Since this is true for any mode $\hat{\mathbf{m}}$, no improvement over the “centered” machine of Sec. IV is possible.

B. Proof of the Symmetry Property

The result (5.1) can be established as follows. Let V in (2.4) be the isometry which gives rise to an ellipsoid $[\mathbf{b}, \boldsymbol{\delta}^B]$, and define a second isometry

$$V'|\alpha\rangle = U(|\alpha\rangle \otimes |\epsilon\rangle) = \sum_{l=0}^3 (\sigma_l|\alpha\rangle) \otimes |e'_l\rangle, \quad (5.6)$$

where the $|e'_l\rangle$ are obtained from the $|e_l\rangle$ by an anti-linear operation (“time reversal”) as explained below.

Choose a reference orthonormal basis $\{|r_j\rangle\}$ for \mathcal{E} , and for any $|e\rangle$ in \mathcal{E} , define its “complex conjugate” $|e^*\rangle$ as the vector with components

$$\langle r_j | e^* \rangle = \langle r_j | e \rangle^*. \quad (5.7)$$

Likewise, if F is any operator on \mathcal{E} , the operator F^* is defined by the matrix elements

$$\langle r_j | F^* | r_k \rangle = \langle r_j | F | r_k \rangle^*, \quad (5.8)$$

that is, the matrix of F^* in this basis is the complex conjugate of that of F . Of course, both $|e^*\rangle$ and F^* depend upon which reference basis is used, but that is immaterial as long as the latter is held fixed throughout the following argument. For later reference, note the relationships

$$\langle e^* | f^* \rangle = \langle e | f \rangle^*, \quad \text{Tr}[F^*] = (\text{Tr}[F])^*. \quad (5.9)$$

Now define the $|e'_l\rangle$ introduced in (5.6) by

$$|e'_0\rangle = |e_0^*\rangle; \quad |e'_q\rangle = -|e_q^*\rangle \text{ for } q = 1, 2, 3. \quad (5.10)$$

It is easy to check that (2.8) and (2.9) are satisfied by the matrix elements

$$E'_{lm} = \langle e'_l | e'_m \rangle \quad (5.11)$$

if they are satisfied by the E_{lm} , so V' with the $|e'_l\rangle$ defined in (5.10) is, indeed, an isometry. The corresponding superoperator \hat{V}' , see (2.11) to (2.13), can be written as

$$\hat{V}'(\sigma_l) = \sum_m \sigma_m \otimes \hat{F}'_{lm} \quad (5.12)$$

with

$$\hat{F}'_{lm} = \sum_{jk} L(jk; lm) |e'_j\rangle \langle e'_k|. \quad (5.13)$$

It is then a straightforward exercise to show, using the properties of $L(jk; lm)$ given in App. A, that

$$\hat{F}'_{lm} = \pm \hat{F}_{lm}^*, \quad (5.14)$$

where the sign is $-$ for $l = 0, m > 0$ and for $l > 0, m = 0$; and is $+$ in all other cases. Consequently, the elements of the B' matrix corresponding to V' , see (2.19), are given by

$$B'_{lm} = \text{Tr}[\hat{F}'_{lm}] = \pm B_{lm} \quad (5.15)$$

with the sign following the same rule as in (5.14); note that the complex conjugate can be ignored when we use the trace formula in (5.9), because the B_{lm} are real. Therefore B' is the same as B , except that

$$B'_{0q} = -B_{0q} \text{ for } q = 1, 2, 3; \quad (5.16)$$

we need not concern ourselves with the B_{q0} , because they vanish. That is, B' is obtained from B by reversing the sign of δ^B , precisely what we need in order to investigate (5.1).

The final step in proving (5.1) is to use (3.13) for both V' and V to show that

$$\Omega'_{\mathcal{E}}(\hat{\mathbf{m}}) = \sum_{q=1}^3 m_q \hat{F}'_{0q} = - \sum_{q=1}^3 m_q \hat{F}_{0q}^* = -\Omega^*_{\mathcal{E}}(\hat{\mathbf{m}}). \quad (5.17)$$

Since $\Omega_{\mathcal{E}}(\hat{\mathbf{m}})$ is a Hermitian operator, its eigenvalues are the same as those of $\Omega^*_{\mathcal{E}}(\hat{\mathbf{m}})$, and since the quality factor is the sum of the absolute values of the eigenvalues, we conclude that

$$\mathcal{Q}'_{\mathcal{E}}(\hat{\mathbf{m}}) = \text{Tr}(|\Omega'_{\mathcal{E}}(\hat{\mathbf{m}})|) = \text{Tr}(|\Omega_{\mathcal{E}}(\hat{\mathbf{m}})|) = \mathcal{Q}_{\mathcal{E}}(\hat{\mathbf{m}}), \quad (5.18)$$

which is (5.1).

C. Proof of the Concavity Property

To establish (5.2) we proceed as follows. Let V^1 and V^2 be any two isometries of the type (2.4):

$$r = 1, 2 : \quad V^r |\alpha\rangle = \sum_{l=0}^3 (\sigma_l |\alpha\rangle) \otimes |e_l^r\rangle. \quad (5.19)$$

Since there is no constraint upon the dimension of the space \mathcal{E} , we can assume that the vectors $|e_l^r\rangle$ are chosen in such a way that

$$\langle e_l^1 | e_m^2 \rangle = 0 \quad (5.20)$$

for all l and m ; in other words, the subspace of \mathcal{E} containing the vectors $\{|e_l^1\rangle\}$ is orthogonal to that containing the vectors $\{|e_l^2\rangle\}$. This, of course, places no restriction upon the matrices

$$E_{lm}^r = \langle e_l^r | e_m^r \rangle, \quad (5.21)$$

aside from the isometry conditions (2.8) and (2.9), which must be satisfied separately for $r = 1$ and 2.

Now construct a third isometry

$$V = \sqrt{p_1} V^1 + \sqrt{p_2} V^2, \quad (5.22)$$

where p_1 and p_2 are non-negative numbers whose sum is 1. That this V is an isometry satisfying (2.7) is a consequence of the fact that both V^1 and V^2 are isometries, along with (5.20), which implies that $(V^1)^\dagger V^2$ and $(V^2)^\dagger V^1$ both vanish. The superoperator corresponding to (5.22) takes the form

$$\hat{V}(\kappa) = V \kappa V^\dagger = p_1 \hat{V}^1(\kappa) + p_2 \hat{V}^2(\kappa) + \sqrt{p_1 p_2} \hat{W}(\kappa), \quad (5.23)$$

where \hat{V}^r is the superoperator corresponding to V^r , and

$$\hat{W}(\kappa) = V^1 \kappa (V^2)^\dagger + V^2 \kappa (V^1)^\dagger \quad (5.24)$$

has the property that

$$\text{Tr}_{\mathcal{E}}[\hat{W}(\kappa)] = 0. \quad (5.25)$$

To see that (5.25) is the case, note that (5.20) implies that we can choose an orthonormal basis $\{|\epsilon_k^r\rangle\}$ for \mathcal{E} , with $r = 1$ or 2, and k taking on as many values as necessary, such that

$$\langle e_l^1 | \epsilon_k^2 \rangle = 0 = \langle e_l^2 | \epsilon_k^1 \rangle \quad (5.26)$$

whatever the values of l and k . Evaluating the trace (5.25) in this basis shows that it is zero.

Hence the partial trace of (5.23) over \mathcal{E} yields

$$\hat{V}_{\mathcal{B}}(\kappa) = p_1 \hat{V}_{\mathcal{B}}^1(\kappa) + p_2 \hat{V}_{\mathcal{B}}^2(\kappa), \quad (5.27)$$

and, as a consequence of (2.18), the corresponding 4×4 B matrices of Sec. II are related by

$$B = p_1 B^1 + p_2 B^2. \quad (5.28)$$

Since B^1 and B^2 can be any two B matrices which are physically possible—the isometries V^1 and V^2 are arbitrary—and p_1 and p_2 any positive numbers whose sum is one, (5.28) shows that the set of physically possible B matrices generated by isometries of the type (2.4) is convex.

Tracing (5.23) over \mathcal{B} yields

$$\hat{V}_{\mathcal{E}}(\kappa) = p_1 \hat{V}_{\mathcal{E}}^1(\kappa) + p_2 \hat{V}_{\mathcal{E}}^2(\kappa) + \sqrt{p_1 p_2} \hat{W}_{\mathcal{E}}(\kappa), \quad (5.29)$$

and $\hat{W}_{\mathcal{E}}(\kappa)$ does not (in general) vanish. If we write the identity on \mathcal{E} in the form

$$I_{\mathcal{E}} = P^1 + P^2 \quad (5.30)$$

where

$$P^r = \sum_j |\epsilon_j^r\rangle\langle\epsilon_j^r| \quad (5.31)$$

is a projector onto the subspace containing the $\{|\epsilon_l^r\rangle\}$, then

$$P^1 \hat{V}_{\mathcal{E}}(\kappa) P^1 = p_1 \hat{V}_{\mathcal{E}}^1(\kappa), \quad P^2 \hat{V}_{\mathcal{E}}(\kappa) P^2 = p_2 \hat{V}_{\mathcal{E}}^2(\kappa), \quad (5.32)$$

and theorem 2 in App. B tell us that

$$\text{Tr}(|\hat{V}_{\mathcal{E}}(\kappa)|) \geq p_1 \text{Tr}(|\hat{V}_{\mathcal{E}}^1(\kappa)|) + p_2 \text{Tr}(|\hat{V}_{\mathcal{E}}^2(\kappa)|). \quad (5.33)$$

Substituting $\kappa = \frac{1}{2}\hat{\mathbf{m}} \cdot \boldsymbol{\sigma}$ into this expression yields (5.2).

VI. OPTIMAL PAIRS AND OPTIMIZATION MAP

A. General Conditions

In Sec. IV we showed that if \mathbf{b} is the three vector of principal axes for the ellipsoid of the \mathcal{B} channel, there is a largest ellipsoid for the \mathcal{C} channel, characterized by principal axes \mathbf{c} and given by the function $G(\mathbf{b})$, (4.1), which represents the best possible copy in \mathcal{C} if \mathbf{b} is held fixed. Note that the components of $\mathbf{c} = G(\mathbf{b})$ are all positive, whereas those of \mathbf{b} may be either positive or negative. In this section we examine the optimization function $G(\mathbf{b})$ in a bit more detail, and deduce necessary and sufficient conditions for an optimal pair (\mathbf{b}, \mathbf{c}) of copies.

If $\mathbf{f} = (f_1, f_2, f_3)$ is a three vector with real components, we define its absolute value to be

$$\mathbf{f}^+ = (|f_1|, |f_2|, |f_3|). \quad (6.1)$$

We shall write $\mathbf{f} \geq 0$ provided each f_p , $1 \leq p \leq 3$, is non-negative, and $\mathbf{f} > 0$ if, in addition, at least one f_p is strictly positive. The relationship $\mathbf{f} \geq \mathbf{g}$ means $\mathbf{f} - \mathbf{g} \geq 0$, or $f_p \geq g_p$ for $1 \leq p \leq 3$, whereas $\mathbf{f} > \mathbf{g}$ means $\mathbf{f} - \mathbf{g} > 0$. The same notation will be used for a four vector such as $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$; β^+ is its absolute value in the sense of (6.1), and $\beta > 0$ means that at least one component is strictly positive and the rest are non-negative.

A three vector \mathbf{b} will be said to be *possible* provided it satisfies (4.10) and is thus a possible collection of semi-axes (some of which may be negative) for the ellipsoid in \mathcal{B} , while a *possible pair* (\mathbf{b}, \mathbf{c}) is one that corresponds to some conceivable copying machine, i.e., is allowed by the laws of quantum mechanics. The pair (\mathbf{b}, \mathbf{c}) will be said to be *better* (in quality) than (\mathbf{b}, \mathbf{c}) provided

$$\mathbf{b}^+ \leq \check{\mathbf{b}}^+, \quad \mathbf{c}^+ \leq \check{\mathbf{c}}^+, \quad (6.2)$$

and at least one of these inequalities is strict, e.g., $\mathbf{c}^+ < \check{\mathbf{c}}^+$. If both inequalities are equalities, the pairs are *equivalent* (in quality). We use absolute values in (6.2) because they alone enter the quality function, see (3.12). Finally, an *optimal pair* (\mathbf{b}, \mathbf{c}) is one for which no other *possible* pair can be found which is better. Note that if (\mathbf{b}, \mathbf{c}) is optimal, so also is (\mathbf{c}, \mathbf{b}) , for it obviously does not matter which output channel is labeled \mathcal{B} and which is labeled \mathcal{C} .

For any possible \mathbf{b} , the argument in Sec. IV, supplemented by the results in Sec. V, tells us that $(\mathbf{b}, G(\mathbf{b}))$ is a possible pair, and for any other possible pair of the form $(\mathbf{b}, \check{\mathbf{c}})$ it is the case that

$$\check{\mathbf{c}}^+ \leq G(\mathbf{b}); \quad (6.3)$$

recall that, by construction, $G(\mathbf{b}) \geq 0$, so it is not necessary to take the absolute value on the right side. This does *not* mean that given any possible \mathbf{b} , $(\mathbf{b}, G(\mathbf{b}))$ is an optimal pair in the sense defined above, because there might very well be a positive optimal pair $(\mathbf{b}', \mathbf{c}')$ with $\mathbf{b}'^+ < \mathbf{b}^+$ and $G(\mathbf{b}) \leq \mathbf{c}'$. Indeed, this can occur even in cases in which $\mathbf{b} = \mathbf{b}^+$ has all positive components. However, if (\mathbf{b}, \mathbf{c}) is an optimal pair, then so is $(\mathbf{b}, G(\mathbf{b}))$, and

$$\mathbf{c}^+ = G(\mathbf{b}). \quad (6.4)$$

The reason is that the pair $(\mathbf{b}, G(\mathbf{b}))$ is possible, and were it better than (\mathbf{b}, \mathbf{c}) , that is, were (6.4) the inequality allowed by (6.3), then (\mathbf{b}, \mathbf{c}) would not be an optimal pair. Thus we see that if (\mathbf{b}, \mathbf{c}) is optimal, so is $(\mathbf{b}, \mathbf{c}^+)$, and, by the same sort of argument, $(\mathbf{b}^+, \mathbf{c})$ and $(\mathbf{b}^+, \mathbf{c}^+)$ are also optimal pairs.

An optimal pair (\mathbf{b}, \mathbf{c}) with $\mathbf{b} \geq 0$ and $\mathbf{c} \geq 0$ will be called a *positive optimal pair*. Note that if we reverse the signs of some of the components of \mathbf{b} or \mathbf{c} , or both, the resulting pair will also be optimal *provided it is possible*. Thus the task of characterizing optimal pairs can be divided into two parts: finding necessary and sufficient conditions for positive optimal pairs, and conditions under which reversing some of the signs in a positive optimal pair results in a possible, and therefore optimal pair.

The optimization map $G(\mathbf{b})$ described in Sec. IV is obtained through a series of steps which can be represented schematically in the form:

$$b \longleftrightarrow \beta^2 \longleftrightarrow \beta \longleftrightarrow \gamma \longrightarrow \gamma^2 \longleftrightarrow c \quad (6.5)$$

in terms of the four-vectors introduced in Sec. IV; see, e.g., (4.5) and (4.16). Here \longleftrightarrow stands for the reversible transformation obtained by multiplying a four vector by the matrix Λ and perhaps a positive constant: see the explicit expressions in (4.6), (4.14), and (4.16). The single arrow means taking the square of each component; thus β^2 is $(\beta_0^2, \beta_1^2, \beta_2^2, \beta_3^2)$. This operation is not reversible because it is many to one or, equivalently, because the square root can be either positive or negative. (Note that $\beta^2 > 0$ is a consequence of the positivity of the matrix E_{lm} , see (2.6) and (4.2), so we do not have to worry about imaginary roots.) Thus in going from the left side to the right side of (6.5) we have to specify the square root of β^2 . The map $G(\mathbf{b})$ defined in Sec. IV employs the positive root, $\beta > 0$.

As a consequence, whereas one or more of the components of \mathbf{b} may be negative (recall that $b_0 = 1 = c_0$), the components of \mathbf{c} , which can be expressed in terms of those of β using (4.17), are always non-negative: $G(\mathbf{b}) \geq 0$. On the other hand, $\beta > 0$ does not imply that the components of γ in (6.5) are positive; indeed, the counterpart of (4.17),

$$b_q = 2(\gamma_0\gamma_q + \gamma_{q'}\gamma_{q''}) \quad (6.6)$$

for $1 \leq q \leq 3$, shows that if some b_q is negative, some of the γ_j must be negative as well.

If in (6.6) we replace every component of γ by its absolute value, the result will be a vector \mathbf{b}' with, obviously,

$$\mathbf{b}' \leq \mathbf{b} = G(\mathbf{c}), \quad (6.7)$$

since $G(\mathbf{c})$ results from following the chain in (6.5) from right to left with γ replaced by the positive square root of γ^2 .

By examining (6.6) for each p in turn, and noting that γ_0 must be strictly positive (half the sum of the β_j), it is easy to show that a necessary and sufficient condition for equality between \mathbf{b}^+ and \mathbf{b}' in (6.7) is that

$$\gamma_4 := \gamma_0\gamma_1\gamma_2\gamma_3 \geq 0. \quad (6.8)$$

Consequently, if γ_4 is negative, (\mathbf{b}, \mathbf{c}) with $\mathbf{c} = G(\mathbf{b})$ cannot be an optimal pair, since $(\mathbf{b}', \mathbf{c})$ with $\mathbf{b}' = G(\mathbf{c})$ is possible, and better. In addition,

$$\begin{aligned} b_1 b_2 b_3 / 8 &= \gamma_0^2 \gamma_1^2 \gamma_2^2 + \gamma_0^2 \gamma_1^2 \gamma_3^2 + \gamma_0^2 \gamma_2^2 \gamma_3^2 \\ &\quad + \gamma_1^2 \gamma_2^2 \gamma_3^2 + \gamma_4(\gamma_0^2 + \gamma_1^2 + \gamma_2^2 + \gamma_3^2), \end{aligned} \quad (6.9)$$

a consequence of (6.6), tells us that if (\mathbf{b}, \mathbf{c}) is an optimal pair, then

$$b_1 b_2 b_3 \geq 0. \quad (6.10)$$

The reason is that if (6.10) is violated, so is (6.8), which would mean that (6.7) is a strict inequality, and $(G(\mathbf{c}), \mathbf{c})$ is better than (\mathbf{b}, \mathbf{c}) . But this is not possible if (\mathbf{b}, \mathbf{c}) is optimal. Note that (6.8) is stronger than (6.10), so even if (6.10) is satisfied, $(\mathbf{b}, G(\mathbf{b}))$ need not be optimal. Of course, (6.10) with b replaced by c must also hold for any optimal pair (\mathbf{b}, \mathbf{c}) .

However, (6.10) is enough to settle the question of whether, given a *positive* optimal pair (\mathbf{b}, \mathbf{c}) , a pair $(\mathbf{b}', \mathbf{c}')$ obtained by reversing the signs of one or more components of \mathbf{b} or \mathbf{c} (or both) is possible, and hence optimal. The answer is that \mathbf{b}' is either \mathbf{b} itself, or obtained from \mathbf{b} by reversing the signs of *precisely two*

components of \mathbf{b} , leaving the sign of the other component fixed. If \mathbf{b} has a component equal to zero, this can, but need not be counted among the two whose sign is reversed, and thus any or all of the non-zero components can be changed in sign without violating (6.10) or the two component rule just mentioned. The same relationship must hold between \mathbf{c}' and \mathbf{c} . Now since reversing the sign of two of the components of \mathbf{b} can be achieved using a suitable unitary transformation on \mathcal{B} , see the discussion in the paragraph preceding (2.28), and a similar comment applies to \mathbf{c} , a $(\mathbf{b}', \mathbf{c}')$ obtained in this way is surely possible, and therefore optimal. On the other hand, reversing the sign of just one component of \mathbf{b} , assuming all three components are non-zero, will violate (6.10), and thus the resulting pair is not possible, since if it were possible it would be optimal. And the same comment applies to \mathbf{c} when its three components are all positive.

In summary, the task of identifying all optimal pairs (\mathbf{b}, \mathbf{c}) reduces to that of identifying all positive optimal pairs. All others are obtained from these by changing the signs of two components of \mathbf{b} and/or two components of \mathbf{c} , since this operation always yields an optimal pair, and any optimal pair which is non-positive is the result of such an operation. In addition, if (\mathbf{b}, \mathbf{c}) is a positive optimal pair, it follows from the definition of optimality and the properties of the map G , see (6.4), that

$$\mathbf{c} = G(\mathbf{b}), \quad \mathbf{b} = G(\mathbf{c}), \quad (6.11)$$

that is, the chain (6.5) will work in either direction, and, therefore,

$$\beta > 0, \quad \gamma = \frac{1}{2}\Lambda \cdot \beta > 0. \quad (6.12)$$

However (6.11), or the equivalent (6.12), while it implies (6.8) and the corresponding condition for β , and is thus stronger than (6.10) or the corresponding condition on \mathbf{c} , is a necessary, but not sufficient condition for (\mathbf{b}, \mathbf{c}) to be a positive optimal pair, because (as noted following (6.3)) there might be a pair $(\mathbf{b}', \mathbf{c}')$ for which $\mathbf{c}' = G(\mathbf{b}')$ and $\mathbf{b}' = G(\mathbf{c}')$, with $\mathbf{b} < \mathbf{b}'$ and $\mathbf{c} < \mathbf{c}'$. To examine this possibility, it is useful to study the Jacobian matrix of the map G . As shown in App. C, one can relate the differentials of \mathbf{b} and \mathbf{c} through

$$d\mathbf{c} = (16\beta_4)^{-1} \mathbf{J} \cdot d\mathbf{b} \quad (6.13)$$

or its inverse

$$d\mathbf{b} = (16\gamma_4)^{-1} \mathbf{K} \cdot d\mathbf{c}, \quad (6.14)$$

where γ_4 is defined in (6.8), and β_4 is the corresponding quantity for β . The 3×3 matrix \mathbf{J} has components

$$J_{qq} = h_{q'} h_{q''}, \quad J_{qq'} = -h_{q'} c_{q''}, \quad J_{q'q} = -h_q c_{q''}, \quad (6.15)$$

while those of \mathbf{K} are

$$K_{qq} = h_{q'} h_{q''}, \quad K_{qq'} = -h_{q'} b_{q''}, \quad K_{q'q} = -h_q b_{q''}, \quad (6.16)$$

and $\mathbf{h} = (h_1, h_2, h_3)$ is a vector whose components are defined by

$$h_q = 2(\beta_0 \beta_q - \beta_{q'} \beta_{q''}) = 2(\gamma_0 \gamma_q - \gamma_{q'} \gamma_{q''}). \quad (6.17)$$

Let us assume that $\gamma_4 > 0$, as will be the case except on the boundaries of the region (6.12); note that this implies, (6.6), that each component of \mathbf{b} is strictly positive. If we can find a $d\mathbf{c}$ whose components are all positive and which, inserted in (6.13), yields a $d\mathbf{b}$ whose components are also positive, then it is clear that (\mathbf{b}, \mathbf{c}) cannot be a positive optimal pair, for there is a better $(\mathbf{b}', \mathbf{c}')$ in its immediate vicinity. It turns out, see App. D, that under the conditions (6.12), at most one of the h_q can be negative. If we suppose that

$$h_1 < 0, \quad h_2 > 0, \quad h_3 > 0, \quad (6.18)$$

the elements K_{q1} of the first column of the \mathbf{K} matrix are all strictly positive, and therefore if $d\mathbf{c} = (1, 0, 0)$ is inserted in (6.14), the resulting $d\mathbf{b}$ will have all three components strictly positive. Clearly the same will be true for a $d\mathbf{c}$ with all positive components which is sufficiently close to the one we have just considered, and, consequently, (6.18), or any other case in which two components of \mathbf{h} are positive and one is negative, is inconsistent with (\mathbf{b}, \mathbf{c}) being a positive optimal pair, as long as $\gamma_4 > 0$.

What this suggests is that in addition to (6.11) or (6.12), we need the condition

$$\mathbf{h} \geq 0, \quad (6.19)$$

in order to ensure that (\mathbf{b}, \mathbf{c}) is a positive optimal pair. Note that we have not established with complete mathematical rigor that (6.19) is a necessary condition for a positive optimal pair, since the argument given above no longer works (at least in any simple form) if, for example, $h_2 > 0$ in (6.18) is replaced by $h_2 = 0$. Such “boundary cases” require additional study, which we have carried through in some simple cases, but not in complete detail. In addition, even if (6.19) is necessary, it does not follow that it is sufficient, in conjunction with (6.11), to guarantee that (\mathbf{b}, \mathbf{c}) is a positive optimal pair, because it refers only to local properties of the G map. While the region defined by (6.19) along with (6.12) is connected, we know of no obvious reason why a given (\mathbf{b}, \mathbf{c}) in this region must be connected to a better $(\check{\mathbf{b}}, \check{\mathbf{c}})$, assuming one exists, through some path along which each component changes monotonically. Such a path can be ruled out by a minor extension of the above analysis as long as everywhere on it all the components of \mathbf{h} , together with γ_4 are strictly positive. But paths along which some components of \mathbf{b} or \mathbf{c} increase while others decrease, followed later by increases and decreases of other components, the final result being an increase in all components, are easy to imagine.

Having not succeeded in supplying an analytical argument, we attacked the problem numerically in the following way. Vectors \mathbf{b} were chosen at random in the “good” region defined by (6.12) and (6.19), and, for each such \mathbf{b} , further vectors $\check{\mathbf{b}}$ lying in the “good” region and also satisfying $\check{\mathbf{b}} > \mathbf{b}$ were chosen at random, and we checked to see whether

$$G(\check{\mathbf{b}}) \geq G(\mathbf{b}). \quad (6.20)$$

With some 4000 choices of \mathbf{b} , and for each of these 100,000 choices of $\check{\mathbf{b}}$, we found no cases in which (6.20) was satisfied, whereas if we made a similar test using \mathbf{b} vectors lying outside the “good” region, a random search turned up many examples of vectors $\check{\mathbf{b}} > \mathbf{b}$ satisfying (6.20).

Consequently, we believe that the necessary and sufficient condition for (\mathbf{b}, \mathbf{c}) to be a positive optimal pair is that \mathbf{b} and \mathbf{c} are both generated using (6.5) from a four vector β satisfying the conditions

$$0 \leq \beta_q \leq \beta_0, \quad \beta_0 \beta_q \geq \beta_{q'} \beta_{q''} \quad (6.21)$$

for $q = 1, 2$ and 3 —note that the second inequality is equivalent to (6.19)—along with the normalization condition (4.8).

Let us define the *class* P to be the collection of all four-vectors which satisfy the two conditions in (6.21) for all q between 1 and 3 , whatever the value of $\sum_j \beta_j^2$. Also, we shall say that any two vectors ξ and η belonging to class P are in (or have) the *same order* provided, for $1 \leq p \leq 3$ and $1 \leq q \leq 3$,

$$\text{sign}(\xi_p - \xi_q) = \text{sign}(\eta_p - \eta_q), \quad (6.22)$$

where $\text{sign}(x)$ is $+1$, 0 or -1 for x positive, zero, or negative. For example, if $\xi_1 \leq \xi_2 \leq \xi_3$, then $\eta_1 \leq \eta_2 \leq \eta_3$, and vice versa; $\xi_1 = \xi_2$ if and only if $\eta_1 = \eta_2$. These definitions lead to the following useful result:

Theorem. If ξ is a four vector in class P , and η another four vector obtained from ξ by one of the three operations (a), (b), and (c) given below, or by the application of any combination of these operations carried out in succession, then η is also in class P , and η has the same order as ξ . The three operations are:

- (a) $\eta = k\xi = (k\xi_0, k\xi_1, k\xi_2, k\xi_3)$ for a constant $k > 0$.
- (b) $\eta = \xi^a = (\xi_0^a, \xi_1^a, \xi_2^a, \xi_3^a)$ for an exponent $a > 0$.
- (c) $\eta = \Lambda \cdot \xi$.

The proof is straightforward, except for (c), for which see App. D.

The theorem is useful when applied to (6.5), because each step on the way from b to c is one of the operations (a), (b) or (c), or a combination of (a) and (c). Consequently, if β satisfies the conditions (6.21), they are also satisfied by γ , and also by b and c . In particular, if β is normalized,

$$\sum_{j=0}^3 \beta_j^2 = 1, \quad (6.23)$$

then for $1 \leq q \leq 3$,

$$0 \leq b_q \leq b_0 = 1, \quad b_q \geq b_{q'} b_{q''}. \quad (6.24)$$

Note that (6.24) is not only implied by (6.21) together with the normalization (6.23), it also implies (6.21) and (6.23), and also the set of conditions for $c = (1, \mathbf{c})$ which correspond to (6.24).

Indeed, (6.24) is a *necessary and sufficient* condition for a triple of numbers $\mathbf{b} = (b_1, b_2, b_3)$ to belong to a positive optimal pair, assuming that the latter can be characterized by (6.12) and (6.19). The reason is that (6.24) tells us that $b = (1, \mathbf{b})$ belongs to class P , and therefore the four vector β^2 belongs to class P and has non-negative components. This, together with the normalization (6.23) corresponding to $b_0 = 1$ in (6.24) means that \mathbf{b} is physically possible. Hence the pair $(\mathbf{b}, G(\mathbf{b}))$ is also possible, it is obviously positive, and it is optimal, since (6.21), thus (6.19), is satisfied.

In addition, the theorem tells us that if (6.24) holds and the semi-axes are in the order

$$b_1 \leq b_2 \leq b_3, \quad (6.25)$$

then the β_q, γ_q are in the same order, and likewise

$$c_1 \leq c_2 \leq c_3, \quad (6.26)$$

for $\mathbf{c} = G(\mathbf{b})$, the other member of the optimal pair. In addition, if two of the \mathcal{B} semi-axes are equal to each other, say $b_1 = b_2$, then the corresponding \mathcal{C} semi-axes are also equal, $c_1 = c_2$. Note, however, that there is no reason to expect b_1 to be equal to c_1 ; one ellipsoid can be small and the other large.

In summary, a necessary and sufficient condition that (\mathbf{b}, \mathbf{c}) be a positive optimal pair is that one member of the pair satisfy (6.24) and the other member be obtained by applying the optimization map G to the first. The second member will then also satisfy the conditions (6.24), and \mathbf{b} and \mathbf{c} will have the same order. All optimal pairs which are not positive optimal pairs are obtained from some positive optimal pair by reversing the signs of two components of the first member of the pair, and/or two components of the second member.

B. Particular Examples

Isotropic copies in which

$$b_1 = b_2 = b_3 = r, \quad c_1 = c_2 = c_3 = s, \quad (6.27)$$

that is, the Bloch ellipsoids for \mathcal{B} and \mathcal{C} are spheres of radii r and s , are obtained by letting

$$\beta_1 = \beta_2 = \beta_3 = \sqrt{(1 - \beta_0^2)/3}. \quad (6.28)$$

Using (4.9) and (4.17), one can show that

$$s = \frac{1}{2} \left[1 - r + \sqrt{(1 - r)(1 + 3r)} \right], \quad (6.29)$$

and, by symmetry, r as a function of s is given by precisely the same functional form. The relationship between the two is shown in Fig. 1. Since r is between 0 and 1, the condition (6.24) for optimal pairs is satisfied. In other words, any point that lies outside the region enclosed by this curve and the two axes is prohibited; this is identical to the no-cloning bound in [12].

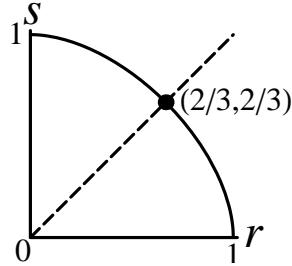


FIG. 1. The relationship between the radii (s and r) of the two output Bloch spheres in the case of isotropic copying. Identical copies occur when $r = s = 2/3$.

As one might expect, there is a certain ‘‘complementarity’’ between the quality of the copies emerging in the \mathcal{B} and \mathcal{C} channels: as r increases, s decreases. Additional insight into the source of this complementarity comes from studying anisotropic situations. Consider, in particular, the extremely anisotropic case in which

$$\beta_1 = \beta_2 = 0, \quad 0 \leq \beta_3 \leq 1/\sqrt{2}; \quad (6.30)$$

the upper bound on β_3 ensures that (6.21) is satisfied. The corresponding \mathbf{b} and \mathbf{c} for optimal copies are given by

$$b_1 = b_2 = 1 - 2\beta_3^2, \quad b_3 = 1, \quad (6.31)$$

$$c_1 = c_2 = 0, \quad c_3 = 2\beta_3\sqrt{1 - \beta_3^2}. \quad (6.32)$$

Note that as β_3 increases, the quality factor c_3 for the third principal mode of the \mathcal{C} copy increases whereas the factor for the same mode of \mathcal{B} remains perfect, $b_3 = 1$. On the other hand, the quality factors for modes 1 and 2 of \mathcal{B} decrease as β_3 increases. Thus improving the quality of a particular mode in one channel leads to a decrease in optimal quality of all the ‘‘perpendicular’’ modes of the other channel. In particular, c_1 and c_2 are 0 (worthless copies) in this case because of the perfect quality for mode 3 of \mathcal{B} .

VII. QUANTUM CIRCUIT

A complicated unitary transformation can often be decomposed into several simpler transformations and expressed in a pictorial fashion by using a quantum circuit. This is helpful both for understanding the transformation and for implementing it in a systematic way. It is worth noting that any unitary transformation can be produced by a circuit which only uses two-qubit XOR gates along with an appropriate collection of one-qubit gates [17].

From Sec. IV, the optimal copying procedure corresponding to $\mathbf{c} = G(\mathbf{b})$ can always be achieved, for any given \mathbf{b} , by using a transformation which generates the isometry (4.13). One example is given by the circuit in Fig. 2(a), where the input state is a tensor product $|\alpha\rangle \otimes |\epsilon\rangle$, with $|\alpha\rangle$ arbitrary, and

$$|\epsilon\rangle = \beta_0|00\rangle + \beta_1|01\rangle + \beta_2|11\rangle + \beta_3|10\rangle. \quad (7.1)$$

Since $|\epsilon\rangle$ involves only real coefficients, it can be produced by a preparation circuit as shown in [10]. The three horizontal lines represent the three qubits going from left to right, as time increases, through four quantum gates. The order of the qubits from top to bottom is the same as the left-to-right order in (4.13).

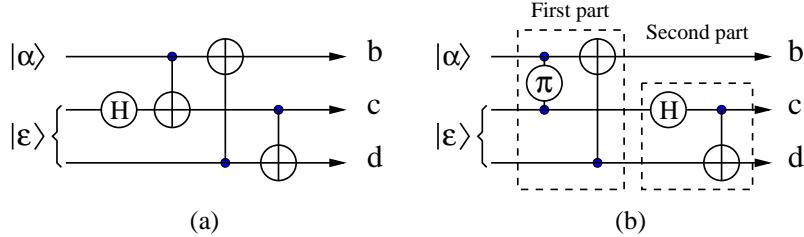


FIG. 2. (a) Optimal copy circuit. The qubit to be copied enters the circuit as $|\alpha\rangle$ on the left, and the copies emerge in b and c on the right. (b) Alternative circuit, see discussion in text.

The first gate in Fig. 2(a) produces the Hadamard transform

$$\begin{aligned} |0\rangle &\mapsto |+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, \\ |1\rangle &\mapsto |-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}, \end{aligned} \quad (7.2)$$

on a single qubit. The remaining three gates are XOR gates acting on different pairs of qubits; each corresponds to the transformation

$$\begin{aligned} |00\rangle &\mapsto |00\rangle, & |10\rangle &\mapsto |11\rangle, \\ |01\rangle &\mapsto |01\rangle, & |11\rangle &\mapsto |10\rangle, \end{aligned} \quad (7.3)$$

in which the left qubit in this formula is the “controlling” and the right qubit the “controlled” qubit. If the controlling qubit is $|0\rangle$, the controlled qubit is left unchanged; if the controlling qubit is $|1\rangle$, the controlled qubit is flipped from $|0\rangle$ to $|1\rangle$, or $|1\rangle$ to $|0\rangle$, which is known as “amplitude-flipping”. In Fig. 2(a) the controlling qubit is indicated by a solid dot, and the corresponding controlled qubit by a plus inside a circle. Thus the controlling qubit is the top, bottom, and middle qubit in the first, second, and third XOR gates.

The transformation produced by the circuit in Fig. 2(a) can be more easily understood if one uses the alternative circuit in Fig. 2(b), which produces exactly the same overall unitary transformation. It employs another gate, indicated by π inside a circle connecting two black dots, which produces the transformation

$$\begin{aligned} |00\rangle &\mapsto |00\rangle, & |10\rangle &\mapsto |10\rangle, \\ |01\rangle &\mapsto |01\rangle, & |11\rangle &\mapsto -|11\rangle, \end{aligned} \quad (7.4)$$

where the phase of $|11\rangle$ is multiplied by a factor of $e^{i\pi} = -1$. It flips one qubit between $|+\rangle$ and $|-\rangle$, see (7.2), if the other qubit is $|1\rangle$. This is known as a “phase flip”, in contrast to the “amplitude flip” produced by an XOR gate. Note that in this case either qubit can be regarded as controlling the other.

Both circuits in Fig. 2 carry out the transformation

$$|\alpha\rangle \otimes |\epsilon\rangle \mapsto \sum_l (\sigma_l |\alpha\rangle) \otimes \beta_l |\hat{\epsilon}_l\rangle \quad (7.5)$$

corresponding to (2.4) combined with (4.3), and thus the isometry (4.13) on the input qubit $|\alpha\rangle$. But Fig. 2(b) is a bit easier to interpret. In the *first part* of the circuit, the lower two qubits control the top qubit: the middle qubit, if it is $|1\rangle$, flips the phase, and the bottom qubit, if $|1\rangle$, flips the amplitude. Thus this part of the circuit results in

$$\begin{aligned} |\alpha\rangle \otimes |00\rangle &\mapsto (\sigma_0 |\alpha\rangle) \otimes |00\rangle, \\ |\alpha\rangle \otimes |01\rangle &\mapsto (\sigma_1 |\alpha\rangle) \otimes |01\rangle, \\ |\alpha\rangle \otimes |11\rangle &\mapsto (\sigma_1 \sigma_3 |\alpha\rangle) \otimes |11\rangle = (-i\sigma_2 |\alpha\rangle) \otimes |11\rangle, \\ |\alpha\rangle \otimes |10\rangle &\mapsto (\sigma_3 |\alpha\rangle) \otimes |10\rangle, \end{aligned} \quad (7.6)$$

where σ_0 is the identity, σ_1 flips the amplitude, σ_3 flips the phase, and $\sigma_1 \sigma_3$ flips the phase and then the amplitude. Consequently, a linear superposition of the maps in (7.6) results in

$$|\alpha\rangle \otimes |\epsilon\rangle \mapsto \sum_l (\sigma_l |\alpha\rangle) \otimes \beta_l |\hat{\epsilon}_l\rangle, \quad (7.7)$$

where

$$|\hat{\epsilon}_0\rangle = |00\rangle, \quad |\hat{\epsilon}_1\rangle = |01\rangle, \quad |\hat{\epsilon}_2\rangle = -i|11\rangle, \quad |\hat{\epsilon}_3\rangle = |10\rangle. \quad (7.8)$$

The *second part* of the circuit in Fig. 2(b) is a unitary transformation on the lower two qubits which maps each $|\hat{\epsilon}_l\rangle$ into $|\hat{\epsilon}_l\rangle$, so that the effect of the two parts combined is (7.5).

This perspective also provides an intuitive meaning for the β_l ’s. The action of the superoperator \hat{V}_B , (2.16), on the input state

$$\alpha = |\alpha\rangle\langle\alpha|, \quad (7.9)$$

given by

$$\hat{V}_B(\alpha) = \sum_l \beta_l^2 (\sigma_l \alpha \sigma_l), \quad (7.10)$$

is determined by the first part of the circuit in Fig. 2(b), thus by (7.7). One can think of (7.10) as a probabilistic mixture of states obtained by applying the unitary transformation σ_l with probability β_l^2 to the input state $|\alpha\rangle$. In other words, β_l^2 for $l \neq 0$ is the probability of producing a noise of type l (amplitude

or phase flipping, or the combination) in the \mathcal{AB} channel. This defines a Pauli channel, in the notation of [12], and since the \mathcal{AC} channel is of the same type—by the symmetry between (4.13) and (4.15)—the corresponding transformation constitutes a Pauli cloning machine as defined in [12]. It is worth mention that even though this Pauli machine is optimal, there exist optimal machines that are not Pauli machines [18].

Finally, note that if $|\epsilon\rangle$ is of the tensor product form

$$\left(\sqrt{1-D_{xy}}|0\rangle + \sqrt{D_{xy}}|1\rangle\right) \otimes \left(\sqrt{1-D_{uv}}|0\rangle + \sqrt{D_{uv}}|1\rangle\right), \quad (7.11)$$

this circuit carries out the optimal eavesdropping in [11] for the case of the BB84 cryptographic system [19]. Here D_{xy} and D_{uv} are the error rates defined in [11]. With $\beta_1 = \beta_2 = \beta_3 = 1/\sqrt{12}$ in (7.1), this circuit carries out Bužek and Hillery's universal cloning. It is an alternative to the circuit given in [10].

VIII. CONCLUSION

We have shown how to produce two optimal copies of one qubit in the cases in which the quality of the copies can be different, and can depend upon which input mode (orthogonal basis) is employed. The measure of quality of the copies is a distinguishability measure on quantum states, and differs from the fidelity measures used in previous published work. In particular, it uses a geometrical representation in which the output possibilities correspond to ellipsoids in the respective Bloch spheres, with a larger ellipsoid corresponding to higher quality. The copy qualities are “complementary” in the sense that improving the quality of one copy for a particular input mode tends to degrade the quality of the other copy for a different set of input modes. Optimal copies can be achieved by means of a relatively simple quantum circuit.

This represents significant progress towards quantifying the no-cloning theorem of quantum information theory. It is useful in applications to quantum cryptography, in which the “copies” go to the eavesdropper and to the legitimate receiver, and also to quantum decoherence processes, in which one “copy” disappears into the environment. In both of these applications, the copies need not be identical, and their qualities can depend upon the input signal.

There are several respects in which one might hope to extend the results presented here. Producing more than two copies of one qubit with (in general) different qualities for the different copies, and qualities which depend upon the input mode remains to be studied. Extending our results to higher-dimensional Hilbert spaces is a significant challenge, because there seems to be no convenient geometrical representations analogous to a Bloch sphere, nor an analytic form of distinguishability measure for more than two states. Nonetheless, recent progress by Zanardi [20] and Cerf [12] in characterizing anisotropic higher-dimensional copies is encouraging. In addition, one could study optimization using other measures of quality or information [21], including Shannon's mutual information. Of course it would be extremely interesting to find some general point of view which unifies all the results on quantum copying obtained up to the present.

ACKNOWLEDGMENTS

One of us (CSN) thanks C. Fuchs for helpful conversations. Financial support for this research has been provided by the NSF and ARPA through grant CCR-9633102.

APPENDIX A: PROPERTIES OF $L(jk; lm)$

The complex conjugate of

$$L(jk; lm) = \frac{1}{2}\text{Tr}[\sigma_j\sigma_l\sigma_k\sigma_m] \quad (A1)$$

is the trace of the σ 's in reverse order. Invariance of the trace under cyclic permutation then shows that

$$L(kj; lm) = L(jk; ml) = L(lm; jk) = L^*(jk; lm). \quad (A2)$$

That is, interchanging j and k , or interchanging l and m , or interchanging jk with lm turns $L(jk; lm)$ into its complex conjugate. Suppose that two 4×4 matrices P and Q are related by

$$P_{jk} = \sum_{lm} L(jk; lm) Q_{lm}. \quad (\text{A3})$$

Then (A2) implies that if Q is real, P is Hermitian, and if Q is Hermitian, P is real.

Explicit values for $L(jk; lm)$ can be obtained in the following way. Suppose that ζ is the number of the four indices j, k, l , and m that are zero. It is obvious that if $\zeta = 4$, $L = 1$, and if $\zeta = 3$, $L = 0$. When $\zeta = 2$, L will vanish unless the non-zero indices are equal, in which case

$$L(00; qq) = L(0q; 0q) = 1. \quad (\text{A4})$$

If $\zeta = 1$, L will vanish unless all the indices are unequal, and by explicit calculation,

$$L(0q; q'q'') = -i, \quad (\text{A5})$$

using the convention (2.10). If $\zeta = 0$, two of the indices must be equal, and L vanishes unless the other two are also equal to each other. So L is zero except for

$$L(qq; qq) = L(qq'; qq') = 1, \quad L(qq; q'q') = -1, \quad (\text{A6})$$

with $q' \neq q$. The other non-zero elements of L can be obtained from (A4)–(A6) by permuting the indices, using (A2).

Regarded as a map, (A3), L carries certain subspaces of the 16-dimensional space of 4×4 matrices into themselves. Thus if $Q_{jk} = 0$ except when jk is equal to 00, 11, 22, and 33, then P_{jk} has the same character, and one can write, see (2.15),

$$P_{jj} = \sum_k \Lambda(j, k) Q_{kk}, \quad (\text{A7})$$

where

$$\Lambda = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (\text{A8})$$

is its own inverse, apart from a factor of 4:

$$\sum_k \Lambda(j, k) \Lambda(k, l) = 4\delta_{jl}. \quad (\text{A9})$$

The other invariant subspaces of L correspond to jk of the form $0q, q0, q'q'', q''q'$ in the notation of (2.10). Hence, thought of as a 16×16 matrix, L is block diagonal with four 4×4 matrices as diagonal blocks. One of these matrices is (A8), and the others can be written down explicitly using (A4) to (A6). In each case the square of the 4×4 matrix is 4 times the identity. Consequently, $\frac{1}{2}L$ is its own inverse, or

$$\sum_{lm} L(jk; lm) L(lm; rs) = 4\delta_{jr}\delta_{ks}. \quad (\text{A10})$$

APPENDIX B: SOME PROPERTIES OF $\text{Tr}(|A|)$

The following results hold for a finite-dimensional Hilbert space \mathcal{H} .

Theorem 1. Let B and C be any two positive (Hermitian with non-negative eigenvalues) operators, and let

$$A = B - C. \quad (\text{B1})$$

Then

$$\mathrm{Tr}(|A|) \leq \mathrm{Tr}(B) + \mathrm{Tr}(C). \quad (\mathrm{B}2)$$

The proof consists in noting that if $\{|\alpha_j\rangle\}$ is an orthonormal basis which diagonalizes A ,

$$A = \sum_j a_j |\alpha_j\rangle\langle\alpha_j|, \quad (\mathrm{B}3)$$

then it follows from (B1) that

$$\begin{aligned} |a_j| &= |\langle\alpha_j|A|\alpha_j\rangle| \leq |\langle\alpha_j|B|\alpha_j\rangle| + |\langle\alpha_j|C|\alpha_j\rangle| \\ &= \langle\alpha_j|B|\alpha_j\rangle + \langle\alpha_j|C|\alpha_j\rangle, \end{aligned} \quad (\mathrm{B}4)$$

where the final equality is a consequence of the fact that diagonal matrix elements of a positive operator cannot be negative. Summing (B4) over j yields (B2)

Theorem 2. Let A be a Hermitian operator, and let

$$I = \sum_k P_k \quad (\mathrm{B}5)$$

be a decomposition of the identity as a sum of mutually orthogonal projectors. Then

$$\sum_k \mathrm{Tr}(|P_k A P_k|) \leq \mathrm{Tr}(|A|). \quad (\mathrm{B}6)$$

In particular, if $\{|\beta_k\rangle\}$ is any orthonormal basis of \mathcal{H} ,

$$\sum_k |\langle\beta_k|A|\beta_k\rangle| \leq \mathrm{Tr}(|A|), \quad (\mathrm{B}7)$$

since we can set $P_k = |\beta_k\rangle\langle\beta_k|$ in (B6).

To prove this theorem, write

$$A = A' - A'', \quad (\mathrm{B}8)$$

with A' that part of the sum (B3) with $a_j \geq 0$, and $-A''$ the remainder. Then A' and A'' are positive operators, and $\mathrm{Tr}(|A|)$ is the sum of their traces. Define

$$A_k = P_k A P_k, \quad A'_k = P_k A' P_k, \quad A''_k = P_k A'' P_k, \quad (\mathrm{B}9)$$

and apply theorem 1 to $A_k = A'_k - A''_k$, to obtain

$$\mathrm{Tr}(|A_k|) \leq \mathrm{Tr}(A'_k) + \mathrm{Tr}(A''_k) = \mathrm{Tr}(A' P_k) + \mathrm{Tr}(A'' P_k), \quad (\mathrm{B}10)$$

where the equality uses $P_k^2 = P_k$. Summing (B10) over k , see (B5), yields (B6).

Theorem 3. Let A be a Hermitian operator on $\mathcal{H} = \mathcal{F} \otimes \mathcal{G}$, and

$$A_{\mathcal{F}} = \mathrm{Tr}_{\mathcal{G}}(A). \quad (\mathrm{B}11)$$

Then

$$\mathrm{Tr}_{\mathcal{F}}(|A_{\mathcal{F}}|) \leq \mathrm{Tr}_{\mathcal{H}}(|A|). \quad (\mathrm{B}12)$$

The proof consists in writing

$$A_{\mathcal{F}} = A'_{\mathcal{F}} - A''_{\mathcal{F}}, \quad (\mathrm{B}13)$$

where $A'_{\mathcal{F}}$ and $A''_{\mathcal{F}}$ are the partial traces over \mathcal{G} of A' and A'' in (B8). As these are positive operators, theorem 1 tells us that

$$\begin{aligned} \mathrm{Tr}_{\mathcal{F}}(|A_{\mathcal{F}}|) &\leq \mathrm{Tr}_{\mathcal{F}}(A'_{\mathcal{F}}) + \mathrm{Tr}_{\mathcal{F}}(A''_{\mathcal{F}}) \\ &= \mathrm{Tr}_{\mathcal{H}}(A') + \mathrm{Tr}_{\mathcal{H}}(A'') = \mathrm{Tr}_{\mathcal{H}}(|A|). \end{aligned} \quad (\mathrm{B}14)$$

APPENDIX C: DIFFERENTIAL OF THE OPTIMIZATION MAP

The differential relationship (6.13) can be obtained using the expressions

$$b_q = \beta_0^2 + \beta_q^2 - \beta_{q'}^2 - \beta_{q''}^2, \quad (C1)$$

$$c_q = 2(\beta_0\beta_q + \beta_{q'}\beta_{q''}), \quad (C2)$$

from (4.9) and (4.17), together with, see (4.8),

$$b_0 = c_0 = \beta_0^2 + \beta_1^2 + \beta_2^2 + \beta_3^2 = 1. \quad (C3)$$

Here and in what follows, $q, q',$ and q'' are related using the convention in (2.10). Let f be some function of $\beta_0, \beta_1, \beta_2, \beta_3$ on the manifold (C3). Its differential can be written as

$$df = \sum_{j=0}^3 (\partial f / \partial \beta_j) d\beta_j = \beta_0^{-1} \sum_{q=1}^3 (\Delta_q f) d\beta_q, \quad (C4)$$

where $d\beta_0$ has been eliminated on the right side by setting the differential of (C3) equal to zero, and

$$\Delta_q f := \beta_0 \partial f / \partial \beta_q - \beta_q \partial f / \partial \beta_0. \quad (C5)$$

Using (C1), (C2) and (C4), one obtains

$$\beta_0 d\mathbf{b} = -4\mathbf{R} \cdot d\beta, \quad (C6)$$

$$\beta_0 d\mathbf{c} = \mathbf{S} \cdot d\beta, \quad (C7)$$

with $d\beta = (d\beta_1, d\beta_2, d\beta_3)$, and the components of the 3×3 matrices \mathbf{R} and \mathbf{S} are given by:

$$R_{qq} = 0, \quad R_{q'q} = R_{q''q} = \beta_q, \quad (C8)$$

$$S_{qq} = 2(\beta_0^2 - \beta_q^2), \quad S_{q'q} = S_{q''q} = c_{q''}. \quad (C9)$$

The inverse of \mathbf{R} is the matrix \mathbf{Q} , with components

$$Q_{qq} = \beta_{q'}\beta_{q''}, \quad Q_{q'q} = Q_{q''q} = -\beta_{q'}\beta_{q''}, \quad (C10)$$

in the sense that

$$\mathbf{Q} \cdot \mathbf{R} = \mathbf{R} \cdot \mathbf{Q} = -2\beta_4 \mathbf{I}, \quad (C11)$$

where $\beta_4 = \beta_0\beta_1\beta_2\beta_3$, and \mathbf{I} is the identity matrix. Multiplying (C6) by \mathbf{Q} , solving for $d\beta$, and inserting the result in (C7) results in (6.13), with \mathbf{J} defined in (6.15), after some straightforward algebra. A symmetrical argument with γ in place of β yields (6.14).

APPENDIX D: MISCELLANEOUS RESULTS FOR SEC. VI

That at most one of the h_q , $q = 1, 2, 3$, can be negative, follows from the condition

$$0 \leq \beta_q \leq \beta_0 \text{ for } q = 1, 2, 3, \quad (D1)$$

which is implied by (6.12), since if the γ_j are non-negative, and one writes out the components of β in terms of those of γ , (4.14), it is obvious that β_0 is at least as large as any of the β_q for $q > 0$. The implications of (D1) can be worked out by assuming, for convenience, the order

$$0 \leq \beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_0. \quad (D2)$$

Then it is obvious that

$$\beta_0\beta_3 \geq \beta_2\beta_1, \quad \beta_0\beta_2 \geq \beta_3\beta_1. \quad (D3)$$

so that, see (6.17), h_2 and h_3 are non-negative, while h_1 might be negative. Similar conclusions follow for other orderings.

The proof of the theorem in Sec. VI for the case $\eta = \Lambda \cdot \xi$ can be constructed using the explicit expressions, see (A8),

$$\begin{aligned}\eta_0 &= \xi_0 + \xi_1 + \xi_2 + \xi_3, \\ \eta_1 &= \xi_0 + \xi_1 - \xi_2 - \xi_3, \\ \eta_2 &= \xi_0 - \xi_1 + \xi_2 - \xi_3, \\ \eta_3 &= \xi_0 - \xi_1 - \xi_2 + \xi_3.\end{aligned}\tag{D4}$$

From these it follows that

$$\eta_q - \eta_{q'} = 2(\xi_q - \xi_{q'}),\tag{D5}$$

in the notational convention of (2.10), and this implies (6.22), which shows that η has the same order as ξ . Next we show that if ξ belongs to class P , that is, its components satisfy the inequalities (6.21), the same is true of η . Because the ξ_i are non-negative, it follows at once from (D4) that

$$\eta_q \leq \eta_0.\tag{D6}$$

Next, a consequence of (D4) is

$$\eta_0\eta_q - \eta_{q'}\eta_{q''} = 4(\xi_0\xi_q - \xi_{q'}\xi_{q''}),\tag{D7}$$

so that if the components of ξ satisfy the second inequality in (6.21), so do those of η . All that remains is to show that

$$\eta_q \geq 0.\tag{D8}$$

If we choose the order of ξ_i , for convenience, to be

$$0 \leq \xi_1 \leq \xi_2 \leq \xi_3 \leq \xi_0,\tag{D9}$$

then η_i has the same order, and it suffices to show that $\eta_1 \geq 0$. Let us assume $\eta_1 < 0$, which by (D4) implies

$$\xi_0 - \xi_2 < \xi_3 - \xi_1.\tag{D10}$$

By (D9), ξ_2 is either zero or positive. If $\xi_2 = 0$, then (D9) implies that $\xi_1 = 0$, and (D10) yields $\xi_0 < \xi_3$, in violation of (D9). If $\xi_2 > 0$, then (D9) and (D10) tell us that

$$\xi_1(\xi_0 - \xi_2) < \xi_2(\xi_3 - \xi_1),\tag{D11}$$

or $\xi_0\xi_1 < \xi_2\xi_3$, which is to say, the components of ξ do not satisfy the second inequality in (6.21). This completes the proof.

- [1] W. K. Wootters and W. H. Zurek, *Nature* **299**, 802 (1982).
- [2] H. Barnum, C. Caves, C. Fuchs, R. Jozsa, and B. Schumacher, *Phys. Rev. Lett.* **76**, 2818 (1996).
- [3] For an introduction to the subject, see C. Bennett, G. Brassard, and A. Ekert, *Scientific American*, Oct. 1992, p. 50, and R. J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. Schauer, *Contemporary Physics* **36**, 149 (1995).
- [4] V. Bužek and M. Hillery, *Phys. Rev. A* **54**, 1844 (1996).
- [5] D. Bruß, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin, *Phys. Rev. A* **57**, 2368 (1998).
- [6] N. Gisin and S. Massar, *Phys. Rev. Lett.* **79**, 2153 (1997).
- [7] D. Bruß, A. Ekert, and C. Macchiavello, [quant-ph/9712019](https://arxiv.org/abs/quant-ph/9712019).

- [8] R. F. Werner, quant-ph/9804001.
- [9] V. Bužek and M. Hillery, quant-ph/9801009.
- [10] V. Bužek, S. L. Braunstein, M. Hillery, and D. Bruß, Phys. Rev. A **56**, 3446 (1997).
- [11] C. Fuchs, N. Gisin, R. B. Griffiths, C.-S. Niu, and A. Peres, Phys. Rev. A **56**, 1163 (1997).
- [12] N. J. Cerf, quant-ph/9805024.
- [13] The same expression derived from a different formalism can be found in I. L. Chuang and M. A. Nielsen, J. of Mod. Opt. **44**, 2455 (1997).
- [14] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991).
- [15] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993) p. 282.
- [16] C. A. Fuchs, quant-ph/9611010.
- [17] A. Barenco *et al.*, Phys. Rev. A **52**, 3457 (1995).
- [18] C.-S. Niu and R. B. Griffiths, in preparation.
- [19] C. Bennett, and G. Brassard, in *Proceedings of the IEEE International Conference on Computer, System, and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
- [20] P. Zanardi, quant-ph/9804011.
- [21] C. A. Fuchs, Ph.D. thesis, University of New Mexico (1995).